THE AMERICAN MATHEMATICAL MONTHLY

Volume 106, Number 10 December 1999

AN OFFICIAL PUBLICATION OF THE MATHEMATICAL ASSOCIATION OF AMERICA

NOTICE TO AUTHORS

The MONTHLY publishes articles, as well as notes and other features, about mathematics and the profession. Its readers span a broad spectrum of mathematical interests, and include professional mathematicians as well as students of mathematics at all collegiate levels. Authors are invited to submit articles and notes that bring interesting mathematical ideas to a wide audience of MONTHLY readers.

The MONTHLY's readers expect a high standard of exposition; they expect articles to inform, stimulate, challenge, enlighten, and even entertain. MONTHLY articles are meant to be read, enjoyed, and discussed, rather than just archived. Articles may be expositions of old or new results, historical or biographical essays, speculations or definitive treatments, broad developments, or explorations of a single application. Novelty and generality are far less important than clarity of exposition and broad appeal. Appropriate figures, diagrams, and photographs are encouraged.

Notes are short, sharply focussed, and possibly informal. They are often gems that provide a new proof of an old theorem, a novel presentation of a familiar theme, or a lively discussion of a single issue.

Articles and Notes should be sent to the Editor:

ROGER A. HORN
1495 E. 100 S., Room 142
University of Utah
Salt Lake City, UT 84112-1114

Please send your email address and 3 copies of the complete manuscript (including all figures with captions and lettering), typewritten on only one side of the paper. In addition, send one original copy of all figures without lettering, drawn carefully in black ink on separate sheets of paper. Authors who use LaTeX are urged to use article.sty and its standard environments with no custom formatting

Letters to the Editor on any topic are invited; please send to the MONTHLY's Utah office. Comments, criticisms, and suggestions for making the MONTHLY more lively, entertaining, and informative are welcome.

See the MONTHLY section of MAA Online for current information such as contents of issues and descriptive summaries of forthcoming articles:

http://www.maa.org/

Proposed problems or solutions should be sent to:

DANIEL ULLMAN, MONTHLY Problems
Department of Mathematics
The George Washington University
2201 G Street, NW, Room 428A
Washington, DC 20052

Please send 2 copies of all problems/solutions material, typewritten on only one side of the paper.

Reprint permission:
DONALD ALBERS, Director of Publications

Advertising Correspondence:
Dave Riska, *driska@maa.org*

Change of address, missing issues inquiries, and other subscription correspondence:
MAA Service Center, *maahq@maa.org*

All at the address:

The Mathematical Association of America
1529 Eighteenth Street, N.W.
Washington, DC 20036

Recent copies of the MONTHLY are available for purchase through the MAA Service Center, maahq@maa.org, 1-800-331-1622

Microfilm Editions: University Microfilms International, Serial Bid coordinator, 300 North Zeeb Road, Ann Arbor, MI 48106.

The AMERICAN MATHEMATICAL MONTHLY (ISSN 0002-9890) is published monthly except bimonthly June-July and August-September by the Mathematical Association of America at 1529 Eighteenth Street, N.W., Washington, DC 20036 and Montpelier, VT. Second class postage paid at Washington, DC, and additional mailing offices. ***Postmaster***: Send address changes to the American Mathematical Monthly, Membership / Subscription Department, MAA, 1529 Eighteenth Street, N.W., Washington, DC, 20036-1385.

Emerging Tools for Experimental Mathematics

Jonathan M. Borwein and Robert M. Corless

1. INTRODUCTION AND WARM-UP

> If I can give an abstract proof of something, I'm reasonably happy. But if I can get a concrete, computational proof and actually produce numbers I'm much happier. I'm rather an addict of doing things on the computer, because that gives you an explicit criterion of what's going on. I have a visual way of thinking, and I'm happy if I can see a picture of what I'm working with.
>
> —John Milnor [**26**, p. 78]

Using mostly elementary examples, we discuss the use of some recent and emerging tools for experimental mathematics. The tools discussed include so-called "inverse symbolic computation", using lattice reduction algorithms such as "LLL" and "PSLQ," and Sloane and Plouffe's integer sequence lookup program. We concentrate on computer-assisted discovery of mathematical results, but a little computer-assisted proof creeps in as well. We use MAPLE throughout the paper, but any other good computer algebra system would be as effective.

This paper is *not* a tutorial on how lattice basis reduction algorithms such as LLL or PSLQ actually work; rather, we discuss some of the ways these tools can be used to generate conjectures, and for that, a detailed understanding of the underlying algorithms is not necessary. We do hope, however, to convey some appreciation of their power.

We begin with some warm-up examples, using the *Inverse Symbolic Calculator* (ISC); http://www.cecm.sfu.ca/MRG/INTERFACES.html. The basic idea is simple: given the first few decimal digits of some real number, we want the ISC to guess a formula for what it 'really' is.

For example, if we input $K_1 = 3.14626436994198$, and click on `simple lookup` (the default) and `Run`, the ISC tells us that

```
...
3146264369941972 = (0405)  1 / abs(- sr(3) + sr(2))
Your value of 314626436994198 would be here.
3146264469611207 = (0192)  (5^(1 / 2) + 4) / (exp(1 / 2) + 1 / 3)
...
```

This has correctly identified K_1 as $1/(\sqrt{3} - \sqrt{2}) = \sqrt{3} + \sqrt{2}$, by table lookup. Using the `integer relation` option would get us, instead, the error message that we need at least 16 digits, and then when we change the final 8 to 72, the following answer appears:

```
K = 3.146264369941972 gave the following results:
      K satisfies the following polynomial, 1 - 10X^2 + X^4
```

together with some negative results about combinations of other constants.

Now consider a second warm-up. If we input the number K_2, computed from the infinite product

$$K_2 = \prod_{n \geq 2} \frac{n^2 - 1}{n^2 + 1} = .2720290549821331\ldots,$$

then the `simple lookup` fails to tell us anything; the `integer relations` option tells us that it is not a simple combination of a few specific constants; but the `smart lookup` tells us that $K_2/2 = \pi/(\exp(-\pi) - \exp(\pi))$. This is actually wrong—it's got the wrong sign, possibly because signs are ignored in this version of the ISC (of course, the program is continually being improved)—but the digits are correctly identified. K_2 is indeed equal to $\pi/\sinh(\pi)$.

As a final warm-up, consider the following two infinite products:

$$K_3 = \prod_{k \geq 1} \frac{\left(1 + \frac{1}{k} + \frac{1}{k^2}\right)^2}{\left(1 + \frac{2}{k} + \frac{3}{k^2}\right)} = 1.84893618285824448\ldots$$

$$K_4 = \prod_{k \geq 1} \frac{\left(1 + \frac{1}{k} + \frac{1}{k^2} + \frac{1}{k^3}\right)^2}{\left(1 + \frac{2}{k} + \frac{3}{k^2} + \frac{4}{k^3}\right)} = 1.797439588835227\ldots$$

Simple lookup, smart lookup, and integer relations as embodied in the ISC all fail to tell us anything about these numbers. In fact, K_3 is

$$K_3 = \prod_{k \geq 1} \frac{\left(1 + \frac{1}{k} + \frac{1}{k^2}\right)^2}{\left(1 + \frac{2}{k} + \frac{3}{k^2}\right)} = \frac{3\sqrt{2}}{\pi} \frac{\cosh^2\left(\pi\frac{\sqrt{3}}{2}\right)}{\sinh(\pi\sqrt{2})},$$

but this is a strange enough formula that we aren't surprised that the ISC can't identify it. We do not know any closed form expression for K_4, however.

The `generalized expansions` option guesses that there is a simple generating function for the 'egyptian fraction' of K_3^{-1}, namely

$$\frac{x(69x - 2)}{47x - 1},$$

but *this is incorrect*, and it is easy to disprove this conjecture by computing the series expansion

$$\frac{x(69x - 2)}{47x - 1} = 2x + \sum_{k \geq 2} (25 \cdot 47^{k-2}) x^k \tag{1}$$

and evaluating the rational number that is the 'egyptian fraction' defined by the coefficients of the series (1):

$$\frac{1}{2} + \sum_{k \geq 2} \frac{1}{25 \cdot 47^{k-2}} = \frac{311}{575} = .540869565\ldots, \tag{2}$$

whereas $K_3^{-1} = 0.5408515498\ldots$, which differs from $311/575$ after the fourth decimal place. Similarly, the ISC's generalized expansions return an incorrect egyptian fraction for K_4^{-1}. Again, note that the ISC is evolving; but some such failures must always be present—its guesses cannot always be correct.

An 'egyptian fraction' is just an ordinary rational written as a sum of reciprocals of natural numbers without repeated entries in the sum.

So, we have seen examples where the ISC tells us something useful, tells us something incorrect, and tells us nothing.

The tools discussed in this paper are only the beginning. The merging of text and tools that can be anticipated over the next few years will make an enormous difference—we can expect greater insight while reading mathematical materials, and easier access to yet more powerful tools—but we make no detailed predictions, because the most significant, qualitative, changes to the work environment are by their nature unexpected. Cases in point are provided by the experiences of the community with MathSci, and with Local Area Networks.

2. A CONNECTION BETWEEN THE LAMBERT W FUNCTION AND STIRLING'S FORMULA FOR $n!$ We now look at a more interesting example, using the online version of the Encyclopedia of Integer Sequences **[28]** (http://www.research.att.com/~njas/sequences/).

The Lambert W function satisfies

$$W(x)e^{W(x)} = x. \tag{3}$$

See **[14]** for a survey of properties and applications of W, together with some of its history; **[16]** explores various series for W, including the one we discuss in this section. We give a short introduction to this function in Appendix A.

There is a branch point of W at $x = -1/e$, where $W(x) = -1$. See Figure 1, a version of which can be produced in MAPLE by the command

```
> plot([t*exp(t),t,t=-5..1],-1..3,-4..1);
```

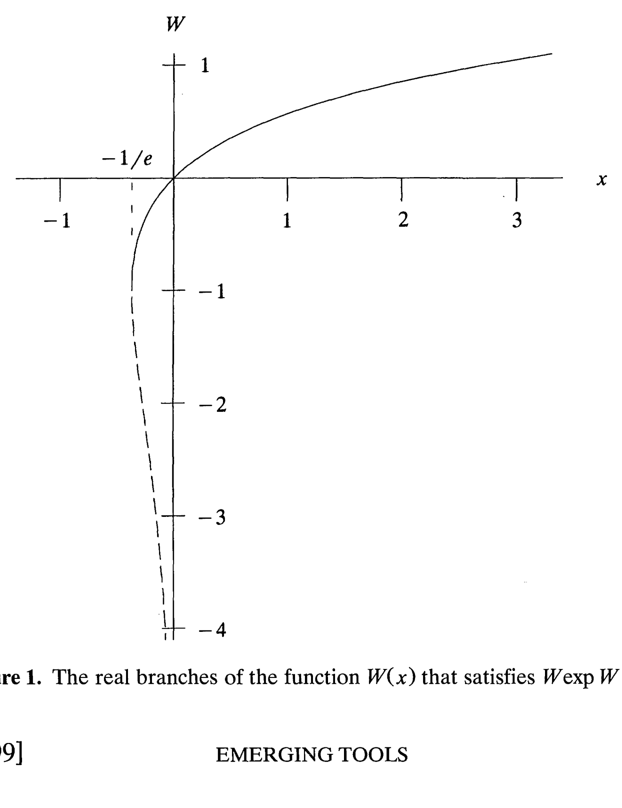


Figure 1. The real branches of the function $W(x)$ that satisfies $W\exp W = x$.

The two real-valued branches of W are denoted $W_0(x)$ and $W_{-1}(x)$; we also refer to $W_0(x)$ as the *principal branch*. We wish to know more about the function near the branch point at $x = -1/e$. After various experiments, we decide to compute the series of

$$W_0(-e^{-1-z^2/2})$$

in MAPLE. We get, very quickly, that

$$W_0(-\exp(-1-z^2/2)) = -1 + z - \frac{1}{3}z^2 + \frac{1}{36}z^3 + \frac{1}{270}z^4 + \frac{1}{4320}z^5 - \frac{1}{17010}z^6 - \frac{139}{5443200}z^7 - \frac{1}{204120}z^8 - \frac{571}{2351462400}z^9 + O(z^{10}). \tag{4}$$

As our first real example of using a new tool, we look up the sequence of denominators 1, 3, 36, 270, 4320, ..., in [**28**]. We find the sequence immediately, and the Encyclopedia gives a reference to the delightful paper [**23**], which does not mention W or refer to any papers on W, or indeed even use it explicitly. Thus, [**23**] would not easily be found by a normal citation search. We find out in [**23**] that equation (4) gives coefficients needed in Stirling's formula for $n!$, which begins

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n}\left(1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \frac{571}{2488320n^4} + O\left(\frac{1}{n^5}\right)\right).$$

The connection we discover (without doing any work ourselves) is that if

$$W_0(-e^{-1-z^2/2}) = \sum_{k\ge 0} (-1)^{k-1} a_k z^k,$$

then

$$n! \sim \sqrt{2\pi n}\, n^n e^{-n} \sum_{k\ge 0} \frac{1\cdot 3\cdot 5\cdots(2k+1)}{n^k} a_{2k+1},$$

and moreover there is a lovely (and useful!) recurrence relation for the a_k's, namely $a_0 = 1$, $a_1 = 1$, and

$$a_n = \frac{1}{(n+1)a_1}\left(a_{n-1} - \sum_{k=2}^{n-1} k a_k a_{n+1-k}\right).$$

3. RIEMANN SURFACES. Tools such as MATLAB and MAPLE permit easy and accurate visualization of Riemann surfaces for elementary functions [**15**], [**29**]. Our qualitative understanding of even extremely basic mathematical building blocks can thus be affected by mathematical software tools. See [**11**] for more discussion of visualization in general; here we concentrate on a simple technique for visualization of Riemann surfaces, namely to make 3-d plots of $\Re f(z)$ or $\Im f(z)$.

It is necessary to *prove* something about this technique—namely, that it really gives us a good picture of the Riemann surface and not just a 3-d plot of the imaginary part (or the real part) of the function involved. This is pursued in more detail in [**15**], but the key point is that given $w = u + iv = f(z) = f(x + iy)$, then we get an accurate Riemann surface by plotting, say, (x, y, v) if and only if the missing piece of information (here, u) is completely determined once x, y, and v are given. This is simple, if not quite obvious: once we have a smooth three-dimensional surface, each point of which can be associated with a unique value (i.e.,

ordered pair) of the map $z \mapsto w = f(z)$, then we have a representation of the Riemann surface of f.

This exact association is not automatic. For example, if $w = \ln(z)$ and we plot (x, y, u), then we do *not* get a picture of the Riemann surface for logarithm, because the branch of $v = \Im(w) = \arg(z)$ is not determined from $u = \ln(x^2 + y^2)/2$, x, and y. If we plot (x, y, v), of course, we *do* recover the classical picture of the Riemann surface for $\ln(z)$, because given x, y, and v we can easily find u.

The following short piece of MAPLE code shows how to graph the Riemann surface for the Lambert W function. We urge you to try the following computation, because the dynamic coloured picture you get is much more easily understood than the static black-and-white image in Figure 2. We also urge you to try your hand at your own functions; many others are graphed in **[15]** and **[29]**.

```
> w := u+I*v;
```

$$w := u + Iv$$

```
> z := w*exp(w);
```

$$z := (u + Iv)e^{(u+Iv)}$$

```
> evalc(z);
```

$$ue^u\cos(v) - ve^u\sin(v) + I(ve^u\cos(v) + ue^u\sin(v))$$

```
> x := evalc(Re(z));
```

$$x := e^u\cos(v) - ve^u\sin(v)$$

```
> y := evalc(Im(z));
```

$$y := ve^u\cos(v) + ue^u\sin(v)$$

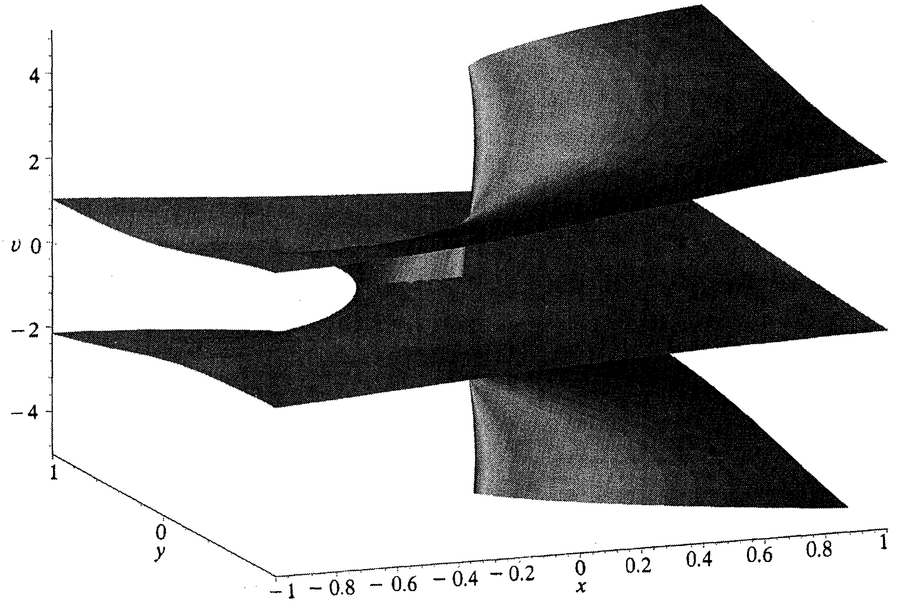


Figure 2. The Riemann surface for the Lambert W function.

```
> plot3d([x,y,v], u=-6..1,
        v=-5..5, axes=FRAME,
        orientation=[-110,73],
        labels=[``x'', ``y'', ``v''],
        style=PATCHNOGRID,
        colour=u,
        view=[-1..1,-1..1,-5..5],
        grid=[50,50]);
```

3.1 One-to-one **correspondence proof.** Given x, y, and v, we have to solve for u. Of course, one takes the existence of (u, v) for a given (x, y) for granted here; for the Lambert W function, a proof can be found in **[14]**. We have

$$(u + iv)e^{u+iv} = x + iy,$$

which gives

$$ue^u + ive^u = (x + iy)e^{-iv} = (x + iy)(\cos v - i \sin v);$$

therefore,

$$ue^u + ive^u = (x \cos v + y \sin v) + i(y \cos v - x \sin v).$$

If $v \neq 0$, and moreover $y \cos v - x \sin v \neq 0$, then dividing the real part by the imaginary part gives u in terms of x, y, and v:

$$u = \frac{v(x \cos v + y \sin v)}{y \cos v - x \sin v}.$$

This solution is unique. Investigation of the exceptional conditions $v = 0$ or $y \cos v - x \sin v = 0$ leads to $u \exp u = x$, which has two solutions if and only if $-1/e \leq x < 0$, in the case $v = 0$, and to the singular condition $u = -\infty$ and $x = y = 0$.

This is precisely what we observe in the graph: two sheets intersect only if $-1/e \leq x < 0$ (note that the colours are different and hence the corresponding sheets on the Riemann surface do not really intersect), and all sheets have a singularity at the origin, except the central one, which contains $v = 0$. This is as good a representation of the Riemann surface for the Lambert W function as can be produced in three dimensions.

However, Figure 2 is nowhere near as intelligible as the live MAPLE plot. On a PC, the use of OpenGL by MAPLE allows the plot to be rotated by direct mouse control. This helps to give a good sense of what the surface is really like, in three dimensions.

4. DYNAMICAL SYSTEMS, NUMERICAL ANALYSIS, AND FORMAL POWER SERIES. In this section we give a brief overview of a surprising connection between numerical analysis of dynamical systems and formal power series. We begin with a simple question: what, exactly, does the fixed time step *forward Euler* numerical method do to the solution of the simple initial value problem

$$\frac{dy}{dt} = y^2 \tag{5}$$

with $y(0) = y_0$? The numerical procedure is just

$$y_{n+1} = y_n + hy'_n \tag{6}$$

for integer $n \geq 0$, where $y'_n = y_n^2$ and $h > 0$ is the chosen time step.

It turns out to be useful to rescale y and t so that $v = hy$ and $\tau = ht$, giving

$$\frac{dv}{d\tau} = v^2, \tag{7}$$

and (6) becomes

$$v_{n+1} = v_n + v_n^2. \tag{8}$$

We may then rephrase our question to ask instead what the relationship between v_n and $v(\tau)$ is.

The point of view taken in **[13]** is that of *backward error analysis*. That is, instead of asking for the difference between $v(n)$ and v_n, we ask instead if there is another differential equation, say

$$\frac{dw}{d\tau} = B(w)w^2, \tag{9}$$

whose solution *interpolates* v_n. That is, we impose the conditions $w(0) = v_0$ and $w(\tau + 1) = w(\tau) + w(\tau)^2$ (*cf.* (8)), and see if we can find such a function $B(w)$.

We do this not so we can improve the behaviour of Euler's method for this problem, but rather so that we may understand what Euler's method has done to the problem; for by understanding the function $B(w)$ we learn something about Euler's method, by comparing (7) to (9).

It turns out that we can use *the method of modified equations* **[19]** to find as many terms of the Taylor series for $B(w)$ as we desire. When we compute the modified equation for (5) to (say) fifth order, we get

$$\frac{dw}{dt} = \left(1 - w + \frac{3}{2!}w^2 - \frac{16}{3!}w^3 + \frac{124}{4!}w^4 - \frac{1256}{5!}w^5\right)w^2. \tag{10}$$

Now we see the sequence $1, -1, 3, -16, 124, -1256$ appearing. This is sequence M3024 in **[28]**, which points us directly to the very beautiful and useful paper **[21]**.

We find in that paper that if

$$B(w) = \sum_{n \geq 0} c_n w^n, \tag{11}$$

then

$$c_n = \frac{1}{n-1} \sum_{i=1}^{n-1} \binom{n-i+1}{i+1} c_{n-i},$$

and this, combined with the functional equation

$$B(w) = \frac{(1+w)^2}{1+2w} B(w + w^2)$$

(which can be iterated to give us two converging infinite products for B), allows us to write an efficient program to evaluate $B(w)$. We can show that $B(w)$ has a pole at $w = -1/2$. By mapping backwards, solving $w + w^2 = -1/2$, we find two more (complex) poles. Iterating this process finds an infinite number of complex poles, approaching the Julia set for the map $v \to v + v^2$ arbitrarily closely; see Figure 3.

The Julia set itself approaches the origin arbitrarily closely. That is, there are poles arbitrarily close to the point of expansion of the series given for B. Thus the series (11) diverges—but, nonetheless, it can be used to evaluate $B(w)$ for w close enough to zero, using MAPLE's built-in sequence acceleration techniques. This is precisely where the convergent infinite products are slow, and hence the series is useful. See **[13]** for details.

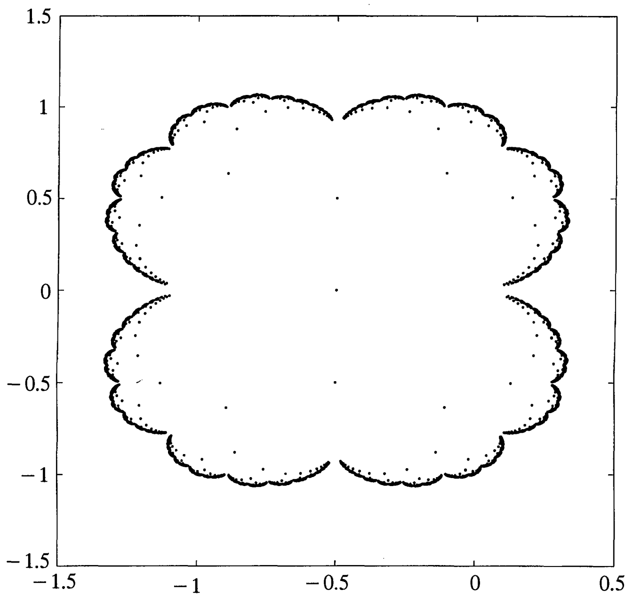


Figure 3. The first 16000 poles of $B(v)$, approaching the Julia set of $v \to v + v^2$.

But more to the point, in [**21**], G. Labelle completely solves the problem of interpolating discrete dynamical systems with continuous dynamical systems, in the domain of formal power series. The mathematical language, however, is quite different from that used in the numerical analysis world. As an example, in [**21**] the 'modified equation' is termed an 'infinitesimal generator' for the discrete dynamical system. Therefore, simple subject searches might not find [**21**]. Indeed, a combinatorics journal seems an unlikely place to find the solution of a problem in the numerical analysis of dynamical systems, but the Encyclopedia of Integer Sequences provides a way to search the 'knowledge database' that is keyed on the *examples*, or the *concrete results*, of papers—not the jargon. This, if you like, is a *new kind* of search tool.

5. AN INTEGER-RELATION EXAMPLE. The following is taken from [**10**]. As a didactic example, suppose that we are interested in finding the value of the definite integral

$$V = \int_0^\infty \frac{\sqrt{x}\ln^5 x}{(1-x)^5}\,dx, \tag{12}$$

and that we suspect that V could be expressed as a polynomial in π^2, of low degree, with short rational coefficients.

Such a conjecture might arise naturally from consideration of

$$\int_0^\infty \frac{\sqrt{x}\ln^2 x}{(1-x)^2}\,dx = 2\pi^2$$

$$\int_0^\infty \frac{\sqrt{x}\ln^3 x}{(1-x)^3}\,dx = \frac{1}{4}\pi^2(\pi^2 - 12)$$

$$\int_0^\infty \frac{\sqrt{x}\ln^4 x}{(1-x)^4}\,dx = -\frac{1}{3}\pi^2(\pi^2 - 12),$$

for example, and we may suppose that these values are known already, for the sake of argument. One can use the `mellin` routine of the `inttrans` package in MAPLE to evaluate all these (and V) symbolically—so this example is really just expository.

To be explicit, we conjecture that

$$V = r_1 + r_2\pi^2 + r_3\pi^6 + r_4\pi^6,$$

where all the r_i are short rational numbers. Instead of trying to derive the coefficients of the this polynomial analytically, we can use numerical approximation and a lattice basis reduction algorithm, the LLL algorithm given in [22], to identify the coefficients heuristically. In an ideal world, we would then know what we had to prove, and, knowing that, would find the proof easier.

We give a short overview of using the LLL algorithm to find integer relations. Suppose that we have a finite set B of n-dimensional linearly independent vectors with rational entries. We call the set

$$L = \left\{ \sum_{v \in B} r_v v \,\middle|\, r_v \in \mathbb{Z} \right\}$$

"the lattice spanned by B." We say that the lattice has dimension n, and that B is a basis for the lattice. There may be many other bases for the lattice, and we often want to find particular bases with nice properties. For many applications, and in particular for finding integer relations, what we would really like to have is "the basis with the shortest Euclidean length." Unfortunately, the problem of determining whether one has *the shortest basis* may be NP-complete [22]. But finding *a short basis* is often just as helpful, and the LLL algorithm [22] can, in polynomial time, find *relatively* short vectors; guaranteed, in fact, to be of length at most $2^{n-1}l$, where l the shortest possible. In practice the LLL algorithm often returns vectors much better than this bound.

To proceed in MAPLE, we choose a large constant C and form the following matrix, and use the lattice reduction subroutine.

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & C \cdot 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & C \cdot \pi^2 \\ 0 & 0 & 1 & 0 & 0 & 0 & C \cdot \pi^4 \\ 0 & 0 & 0 & 1 & 0 & 0 & C \cdot \pi^6 \\ 0 & 0 & 0 & 0 & 1 & 0 & C \cdot \pi^8 \\ 0 & 0 & 0 & 0 & 0 & 1 & C \cdot V \end{bmatrix}$$

```
> readlib( lattice );
```

proc () . . . end

We work to 30 digits for this example. In general, one has to experiment to find how many digits one needs.

```
> Digits := 30;
```

Digits := 30

We compute an approximation for the value V that we wish to identify, and approximations of the quantities that we wish to relate to V.

```
> V := Int( sqrt(x)*ln(x)^5 / (1-x)^5, x=0..infinity);
```

$$V := \int_0^\infty \frac{\sqrt{x}\ln(x)^5}{(1-x)^5}\, dx$$

```
> lastcol := [ seq( evalf( Pi^(2*i)), i=0..4 ), evalf(V)];
```

lastcol := [1., 9.869604401089358618834490999 88,

97.409091034002437236440332688 8, 961.38919357530443703021944365 3,

9488.531016070574007128575503 92, −16.699473719229070496187243400 7]

We now choose a large constant C. We use the size of C to penalize vectors that do not combine to zero.

```
> C := 10^15;
```

$$C := 1000000000000000$$

We construct the rows of the matrix that we need, as follows.

```
> for i to 6 do
>    row.i*j:[ seq(0, j=1..7) ]:
>    row.i[i] := 1:
>    row.i[7] := C* lastcol[i]:
> od:
> B := [ seq( row.i, i=1..6 ) ];
```

$$\begin{aligned} B := \big[&[1,0,0,0,0,0,.1000000000000000\,10^{16}],\\ &[0,1,0,0,0,0,.9869604401089358618834490999988\,10^{16}],\\ &[0,0,1,0,0,0,.9740909103400243723644033268888\,10^{17}],\\ &[0,0,0,1,0,0,.9613891935753044370302194436 53\,10^{18}],\\ &[0,0,0,0,1,0,.9488531016070574007128575503 92\,10^{19}],\\ &[0,0,0,0,0,1,-.1669947371922907049618724340 07\,10^{17}]\big] \end{aligned}$$

Now we call the `lattice` routine to compute a short basis for the set generated by these rows.

```
> lattice( B );
```

$$\begin{aligned} \big[&[0,120,140,-15,0,24,.622\,10^{-11}],\\ &[-16743,51,156,10,-1,-55,6738.90916826007994],\\ &[35146,-443,-57,-16,-1,21,19729.34720281002100],\\ &[6349,-2221,94,2,0,-269,-7554.67120587589348],\\ &[-2452,-99,8,-3,2,805,5948.36266979182662],\\ &[32181,345,9,-11,-1,982,-19383.09100001444674]\big] \end{aligned} \tag{13}$$

All of these new basis vectors are of the form

$$\left[r_1, r_2, r_3, r_4, r_5, r_6, C\sum_{i=1}^{6} r_i a_i\right],$$

where the r_i are integers. This is because each new vector is an integer linear combination of rows of the initial matrix. Because the initial matrix was an augmented identity matrix, the coefficients of the requisite integer combination show up in the result. Because we chose C to be so large, looking for a short vector in this space really biases the search towards places where the integer linear

combination of the final column is zero, if there are any. Hence we suspect, from the first row of (13), that

$$120\pi^2 + 140\pi^4 - 15\pi^6 + 24V = 0$$

or

$$V = \frac{5}{24}\pi^2(3\pi^4 - 28\pi^2 - 24).$$

Issuing the following MAPLE command lends credence to our suspicion.

```
> evalf( V-5/24*Pi^2*(3*Pi^4-28*Pi^2-24), 100);
```

$$-.8\,10^{-97}$$

There is a simpler Web-based implementation, which uses the "EZface" to emulate a more comprehensive GNU MP implementation of this method. Go to http://www.cecm.sfu.ca/MRG/INTERFACES.html, click on `EZface`, and type in the following:

```
lindep([1.,
9.8696044010893586188344909998 8,
97.409091034002437236440332688 8,
961.38919357530443703021944365 3,
9488.5310160705740071285755039 2,
-16.699473719229070496187243400 7])
```

Then, select 30 digits of precision, and click `evaluate`. Very quickly, the vector

```
0,-120.,-140.,15.,0,-24.
```

is returned—voilá, our integer relation.

Issuing the command `lindep` calls a subroutine that looks for short integer linear dependencies among the given vector of numbers. Again its results are to be considered as *possible relations*, to be proved later.

Numerical instability in the LLL algorithm may cause difficulty, as well. Here we have simply worked to enough digits to mitigate its effects—that is, we are trying to buy more accuracy by paying for more precision. This is often expensive, and PSLQ, discussed in Section 6.1, is better, being more stable and hence faster and more reliable.

However, the simple LLL approach is still very powerful and, if used with imagination, offers rich possibilities for discovery.

6. HOW SOLVABLE IS 'SOLVABLE'? This example is also taken from [**10**]. The following problem arises when thinking about modular (theta) functions; see [**6**]. If we define

$$a(q) := \sum_{m,n\in\mathbb{Z}} q^{m^2+mn+n^2}$$

$$b(q) := \sum_{m,n\in\mathbb{Z}} \omega^{n-m} q^{m^2+mn+n^2}$$

$$c(q) := \sum_{m,n\in\mathbb{Z}} q^{(n+1/3)^2+(n+1/3)(m+1/3)+(m+1/3)^2}$$

where $\omega = \exp(2\pi i/3)$, then we have

$$a^3 = b^3 + c^3$$

and a lovely parameterization of the ${}_2F_1$ hypergeometric function [4]:

$$F\left(\begin{array}{c} \frac{1}{3}, \frac{2}{3} \\ 1 \end{array} \middle| \frac{c^3}{a^3}\right) = a.$$

Choosing $q = \exp(-2\pi\sqrt{N/3})$ for $N \in \mathbb{Q}$, it can be shown that $s_N := c/a$ is an algebraic number expressible by radicals; see [6]. If N is a positive integer, then s_N is called the Nth *cubic singular value*. What can we discover computationally about s_N? For example, can we determine radical formulae for the higher order cubic singular values?

The following observations help the efficiency of the computations. It is known that

$$\begin{aligned} a(q) &= \theta_3(q)\theta_3(q^3) + \theta_2(q)\theta_2(q^3) \\ b(q) &= \left(3a(q^3) - a(q)\right)/2 \\ c(q) &= \left(a(q^{1/3}) - a(q)\right)/2, \end{aligned}$$

where

$$\theta_2(q) = \sum_{n \in \mathbb{Z}} q^{(n+1/2)^2} \quad \text{and} \quad \theta_3(q) = \sum_{n \in \mathbb{Z}} q^{n^2}$$

are the classical theta functions. The lacunarity of these series allows for very rapid computation.

6.1 A useful transformation. A further transformation, which makes the as yet unknown minimal polynomial simpler, is useful. After examining the patterns in the first few cases $s_1, s_2, s_3, \ldots$, and using the analogous classical quadratic singular values (where one sees the forms $4k_N^2(1 - k_N^2)$ or $(1 - k_N^2)/2k_N$ depending on the parity of N), the authors of [10] thought to look at

$$G_N := \left(\frac{1}{2} - s_N\right)^2 \qquad N \not\equiv 0 \bmod 3$$

or

$$g_N := \frac{3s_N}{1 - s_N^3} \qquad N \equiv 0 \bmod 3;$$

the minimal polynomial for G_N or g_N then, by observation, has lower degree than the minimum polynomial for s_N. This makes the polynomial easier to find by the PSLQ algorithm.

The PSLQ algorithm (see [17]) and the LLL algorithm can both be used to find integer relations (and hence minimal polynomials for an algebraic number α, by looking for an integer relation among $1, \alpha, \ldots, \alpha^m$). However, PSLQ can also produce *negative* results. If PSLQ fails to find an integer relation, then one can usually say that *there is no such relation with coefficients less than a computable bound*, effectively proving that there is no simple relation of the guessed form.

The authors of [10] used these ideas to 'decode' the numerical values of s_N into radical form, up to $N = 100$, and some values beyond, such as $N = 110$ and 154. They used a variety of strategies to verify the results; some ingenuity was necessary in order to extract the radicals. For $N < 53$, they computed P_N, the minimal polynomial for G_N or g_N; they then tried factoring P_N over different quadratic number fields until they got a factor of degree 4 or less, which they solved in

radicals. This approach failed at $N = 53$, where they first had to use a special MAPLE program for finding a radical for any solvable quintic. (See ftp://calfor.lip6.fr:/pub/softwares/Maple/quinticV2.gz.) The radical returned for $N = 53$ has over 7500 symbols in it. Kevin Hare at the CECM refined it to an equivalent but simpler radical with 'only' 860 symbols. MAPLE was able to verify symbolically that this simpler radical solved P_{53}. In general, determining that a symbolic equation is indeed zero is, in certain classes of expressions, computationally undecidable [**27**].

Indeed, the point of this whole exercise was to determine how good both MAPLE's symbolic tools and PSLQ's numerical ones were on "grand challenge" examples. Experience with exercises such as this have led to improvements in both tools.

Reassurance that the results are correct can often be obtained by using Klein's *absolute invariant* [**3**, p. 115]

$$J_2(x) = \frac{4}{27}\frac{\left(1 - x^2(1 - x^2)\right)^3}{x^4(1 - x^2)^2},$$

and its cubic counterpart

$$J_3(x) = \frac{1}{64}\frac{(9 - 8x^3)^3}{x^9(1 - x^3)}.$$

If our computed s_N is correct, then it is related to the (known) classical singular value k_{3N} by

$$J_2(k_{3N}) = J_3(s_N). \tag{14}$$

The identity (14) can be derived from Proposition 5.8 in [**3**, p. 185, (5.5.26)]. It can be checked symbolically in MAPLE for the radicals arising in the cases $N \le 10$. For larger N, some human intervention is required. For $N = 70$, the verification requires use of k_{210}, the computation of which Hardy called "one of the most striking of Ramanujan's results" [**20**, p. 229]. We note that purely numerical computation, together with analytic reasoning about such computation (some of which is automatable) can be used to verify the results. Standard irrational number theoretic techniques allow one to show that either $J_2(k_{210}) = J_3(S)$ or $|J_2(k_{210}) - J_3(S)| > 10^{-6400}$, where S is our heuristically guessed radical formula for s_{70}. Given this knowledge, a few minutes of CPU time establishes that $|J_2(k_{210}) - J_3(S)| < 10^{-6400}$, and thus $J_2(k_{210}) = J_3(S)$.

7. FINAL VIGNETTES. Integer relation algorithms have already helped to discover many new results. We list a few of these, again taken from [**10**]. The number of such results continues to climb. We have to tell the algorithms what kind of relationship to look for, but, given that, the algorithms allow previously impossible jumps.

7.1 Zeta value series. The formula for $\zeta(3)$ used by Apéry to prove that $\zeta(3)$ is irrational, namely

$$\zeta(3) = \frac{5}{2}\sum_{k \ge 1}\frac{(-1)^{k+1}}{k^3\binom{2k}{k}},$$

has no analog for $\zeta(2n+1)$ with $n \ge 2$; it is not yet known if these ζ values are

irrational. It can be shown using PSLQ (or more simply in this case by the Euclidean algorithm, since there are only two unknown integers) that if a formula like

$$\zeta(5) = \frac{p}{q} \sum_{k \geq 1} \frac{(-1)^{k+1}}{k^5 \binom{2k}{k}}$$

exists, then the integers p and q are larger than 10^{300}.

There is a similar but more complicated formula for $\zeta(5)$, due to Koecher, that *does* suggest generalization, however. Borwein and Bradley used an LLL algorithm to determine the new coefficients [**7**]. They found that

$$\zeta(7) = \frac{5}{2} \sum_{k \geq 1} \frac{(-1)^{k+1}}{k^7 \binom{2k}{k}} + \frac{25}{2} \sum_{k \geq 1} \frac{(-1)^{k+1}}{k^3 \binom{2k}{k}} \sum_{j=1}^{k-1} \frac{1}{j^4},$$

and they discovered similar formulas for $\zeta(4n+3)$ for $2 \leq n \leq 10$ that involve linear combinations of sums of the form

$$H_{m,n} = \sum_{k \geq 1} \frac{(-1)^{k+1}}{k^{4m+3} \binom{2k}{k}} \sum_{j=1}^{k-1} \frac{1}{j^{4(n-m)}}$$

and multiple dimensional analogues. They conjectured the following generating function:

$$\begin{aligned} \sum_{n \geq 0} \zeta(4n+3) z^{4n} &= \sum_{k \geq 1} \frac{1}{k^3(1 - z^4/k^4)} \\ &= \frac{5}{2} \sum_{k \geq 1} \frac{(-1)^{k+1}}{k^3 \binom{2k}{k}} \frac{1}{(1 - z^4/k^4)} \prod_{m=1}^{k-1} \frac{1 + 4z^4/m^4}{1 - z^4/m^4}, \end{aligned} \tag{15}$$

where the final infinite sum is quite unexpected. However, from the first ten cases it was apparent that the series had the form

$$\frac{5}{2} \sum_{k \geq 1} \frac{(-1)^{k+1}}{k^3 \binom{2k}{k}} \frac{1}{(1 - z^4/k^4)} P_k(z)$$

for as yet undetermined P_k; and there were abundant data to compute

$$P_k(z) = \prod_{m=1}^{k-1} \frac{1 + 4z^4/m^4}{1 - z^4/m^4}.$$

They reduced the conjectured formula to an equivalent finite sum

$$\frac{5}{2} \sum_{k=1}^{n} \binom{2k}{k} \frac{k^2}{4n^4 + k^4} \prod_{j=1}^{k-1} \frac{n^4 - j^4}{4n^4 + j^4} = \frac{1}{n^2} \tag{16}$$

($1 \leq n < \infty$) that was subsequently proved by Almkvist and Granville [**1**]. Series expansion of the finite products in (15) gives a rapidly converging series for any $\zeta(4n+3)$. The original motivation for the search for these formulae was the hope that they would shed light on whether these ζ values are irrational.

7.2 Independent computation of digits of π. The following formula, discovered using the PSLQ algorithm, allows rapid computation of hexadecimal digits of π independently of previous digits [**2**]:

$$\pi = \sum_{\geq 0} \left(\frac{1}{16}\right)^k \left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6}\right). \tag{17}$$

Bailey, Borwein, and Plouffe knew that a fast algorithm would result from a formula of this form, and deliberately used a computer search to find it; some have called this approach *mathematical reverse engineering*. Once known, the formula can be proved very concisely by a human [**2**]. Interestingly, the following MAPLE session shows that it can now be proved almost automatically, too.

```
> p := Sum(   (1 / 16)^k* (4 / (8*k + 1)- 2 / (8*k + 4)- 1 / (8*k + 5)
-1 / (8*k+6)),  k=0..infinity );
```

$$p := \sum_{k=0}^{\infty} \left(\frac{1}{16}\right)^k \left(4\frac{1}{8k+1} - 2\frac{1}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6}\right)$$

The following shows a temporary increase in complexity. This phenomenon is called "intermediate expression swell."

```
> value( p );
```

$$\frac{47}{15}\mathrm{hypergeom}\left(\left[1, \tfrac{1}{2}, \tfrac{3}{4}, \tfrac{5}{8}, \tfrac{1}{8}\right], \left[\tfrac{3}{2}, \tfrac{13}{8}, \tfrac{9}{8}, \tfrac{7}{4}\right], \tfrac{1}{16}\right)$$

$$+ \frac{47}{8192} \frac{\mathrm{hypergeom}\left(\left[2, \tfrac{3}{2}, \tfrac{13}{8}, \tfrac{9}{8}, \tfrac{7}{4}\right], \left[\tfrac{5}{2}, \tfrac{21}{8}, \tfrac{17}{8}, \tfrac{11}{4}\right], \tfrac{1}{16}\right)}{\tfrac{123669}{40960} - \tfrac{819}{40960}\sqrt{241}}$$

$$+ \frac{1504}{36855} \frac{\left(\tfrac{391}{8192} - \tfrac{1}{8192}\sqrt{241}\right)\mathrm{hypergeom}\left(\left[2, \tfrac{3}{2}, \tfrac{13}{8}, \tfrac{9}{8}, \tfrac{7}{4}\right], \left[\tfrac{5}{2}, \tfrac{21}{8}, \tfrac{17}{8}, \tfrac{11}{4}\right], \tfrac{1}{16}\right)}{\left(\tfrac{151}{240} + \tfrac{1}{240}\sqrt{241}\right)\left(\tfrac{151}{240} - \tfrac{1}{240}\sqrt{241}\right)}$$

$$+ \frac{47}{1920} \frac{\left(\tfrac{391}{8192} - \tfrac{1}{8192}\sqrt{241}\right)\mathrm{hypergeom}\left(\left[3, \tfrac{5}{2}, \tfrac{21}{8}, \tfrac{17}{8}, \tfrac{11}{4}\right], \left[\tfrac{7}{2}, \tfrac{29}{8}, \tfrac{25}{8}, \tfrac{15}{4}\right], \tfrac{1}{16}\right)}{\left(\tfrac{151}{240} + \tfrac{1}{240}\sqrt{241}\right)\left(\tfrac{151}{240} - \tfrac{1}{240}\sqrt{241}\right)\left(\tfrac{511819}{8192} - \tfrac{1309}{8192}\sqrt{241}\right)}$$

Looking at those conjugate radicals in the denominators suggests expansion—this step is natural but not automatic.

```
> normal( %, expanded );
```

$$\frac{47}{15}\mathrm{hypergeom}\left(\left[1, \tfrac{1}{2}, \tfrac{5}{8}, \tfrac{3}{4}, \tfrac{1}{8}\right], \left[\tfrac{3}{2}, \tfrac{9}{8}, \tfrac{13}{8}, \tfrac{7}{4}\right], \tfrac{1}{16}\right)$$

$$+ \frac{271}{39312}\mathrm{hypergeom}\left(\left[2, \tfrac{3}{2}, \tfrac{9}{8}, \tfrac{13}{8}, \tfrac{7}{4}\right], \left[\tfrac{5}{2}, \tfrac{17}{8}, \tfrac{21}{8}, \tfrac{11}{4}\right], \tfrac{1}{16}\right)$$

$$+ \frac{1}{20944}\mathrm{hypergeom}\left(\left[3, \tfrac{5}{2}, \tfrac{17}{8}, \tfrac{21}{8}, \tfrac{11}{4}\right], \left[\tfrac{7}{2}, \tfrac{25}{8}, \tfrac{29}{8}, \tfrac{15}{4}\right], \tfrac{1}{16}\right) \tag{18}$$

As an aside, equation (18) is an interesting identity itself. In the notation of **[18]**, it implies (once the proof is completed) that

$$\pi = \frac{47}{15} F\left(\begin{matrix} 1, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}, \frac{1}{8} \\ \frac{3}{2}, \frac{9}{8}, \frac{13}{8}, \frac{7}{4} \end{matrix} \,\middle|\, \frac{1}{16} \right) + \frac{271}{39312} F\left(\begin{matrix} 2, \frac{3}{2}, \frac{9}{8}, \frac{13}{8}, \frac{7}{4} \\ \frac{5}{2}, \frac{17}{8}, \frac{21}{8}, \frac{11}{4} \end{matrix} \,\middle|\, \frac{1}{16} \right)$$
$$+ \frac{1}{20944} F\left(\begin{matrix} 3, \frac{5}{2}, \frac{17}{8}, \frac{21}{8}, \frac{11}{4} \\ \frac{7}{2}, \frac{25}{8}, \frac{29}{8}, \frac{15}{4} \end{matrix} \,\middle|\, \frac{1}{16} \right).$$

The next step in the mechanical proof of the Bailey-Borwein-Plouffe π formula simplifies the hypergeometric functions:

```
> convert( %, StandardFunctions );
```

$$-\frac{1}{2}\sqrt{2}\left(\ln\left(1 - \frac{1}{2}\sqrt{2}\right) - \ln\left(1 + \frac{1}{2}\sqrt{2}\right) + \frac{1}{2}\sqrt{2}\ln\left(\frac{1}{2}\right) - \sqrt{2}\arctan(1)\right.$$
$$\left. - 2\arctan\left(\frac{1}{2}\sqrt{2}\right) - \frac{1}{2}\sqrt{2}\ln\left(\frac{5}{2}\right) - \sqrt{2}\arctan\left(\frac{1}{3}\right)\right)$$
$$+ \ln\left(\frac{3}{4}\right) - \ln\left(\frac{5}{4}\right) + \frac{1}{2}\sqrt{2}\left(\ln\left(1 - \frac{1}{2}\sqrt{2}\right) - \ln\left(1 + \frac{1}{2}\sqrt{2}\right) - \frac{1}{2}\sqrt{2}\ln\left(\frac{1}{2}\right)\right.$$
$$\left. + \sqrt{2}\arctan(1) - 2\arctan\left(\frac{1}{2}\sqrt{2}\right) + \frac{1}{2}\sqrt{2}\ln\left(\frac{5}{2}\right) + \sqrt{2}\arctan\left(\frac{1}{3}\right)\right)$$
$$+ \ln\left(\frac{1}{2}\right) - \ln\left(\frac{3}{2}\right) + 2\arctan\left(\frac{1}{2}\right)$$

The next step is not necessary, but it slows down the computation so we can see that many of the terms in the above formula simply cancel.

```
> expand( % );
```

$$\frac{1}{2}\pi + 2\arctan\left(\frac{1}{3}\right) + 2\arctan\left(\frac{1}{2}\right)$$

Now, finally, our answer is plain:

```
> simplify( % );
```

$$\pi$$

A somewhat more efficient version of (17) was discovered by Fabrice Bellard. This has led Colin Percival, an undergraduate student at Simon Fraser, to design an ingenious parallel internet computation of staggeringly high order hexadecimal digits of π. Details may be found at http://www.cecm.sfu.ca/projects/pihex/: the five trillionth bit of π is ‘0’.

7.3 Fast series for the Catalan constant. Consider the *Catalan constant*, which can be defined by

$$G = \sum_{k \geq 0} \frac{(-1)^k}{(2k+1)^2} \tag{19}$$

or alternatively by

$$G = -\int_0^{\pi/4} \log \tan \theta \, d\theta = -\int_0^1 \frac{\log u}{1+u^2}\, du.$$

This is perhaps the simplest constant whose irrationality is still unsettled.

Ramanujan discovered the following series for G, which converges *much* more quickly than (19) [**3**, p. 386]:

$$G = \frac{\pi}{8}\log(2+\sqrt{3}) + \frac{3}{8}\sum_{k\ge 0}\frac{1}{(2k+1)^2\binom{2k}{k}}. \tag{20}$$

After many false starts, David Bradley found a new family of series that includes (20). One member of this family is

$$G = \frac{\pi}{8}\log\left(\frac{10+\sqrt{50-22\sqrt{5}}}{10-\sqrt{50-22\sqrt{5}}}\right) + \frac{5}{8}\sum_{k\ge 0}\frac{L_{2k+1}}{(2k+1)^2\binom{2k}{k}}. \tag{21}$$

where the *Lucas numbers* L_n are given by $L_n = L_{n-1} + L_{n-2}$ with $L_0 = 2$, $L_1 = 1$.

The general formula for Bradley's family of series is proved using certain identities among log tangent integrals. For example, (21) is proved using

$$2\int_0^{\pi/4}\log(\tan\theta)\,d\theta = 5\int_0^{3\pi/20}\log(\tan\theta)\,d\theta - 5\int_0^{\pi/20}\log(\tan\theta)\,d\theta.$$

This identity was discovered by an LLL integer relation algorithm. It turns out to be quite easy to search for such relations among log tangent integrals, whereas looking for resummations of the original series (by LLL) is quite difficult.

David Broadhurst has, in his pursuit of new insights for theoretical physics, computationally probed more of these constants [**12**]. Based on an extraordinary blend of intuition, methodical use of PSLQ, and computer-assisted proofs, he was led to remarkable binary identities for polylogarithmic constants such as $\zeta(3)$, $\zeta(5)$, and Catalan's constant. His formula for Catalan's constant is:

$$\begin{aligned} G = {}& \frac{3}{2}\sum_{i=0}^{\infty}\frac{1}{16^i}\left(\frac{1}{(8i+1)^2} - \frac{1}{(8i+2)^2} + \frac{1}{2}\frac{1}{(8i+3)^2}\right.\\ &\qquad\left. - \frac{1}{4}\frac{1}{(8i+5)^2} + \frac{1}{4}\frac{1}{(8i+6)^2} - \frac{1}{8}\frac{1}{(8i+7)^2}\right)\\ & - \frac{1}{4}\sum_{i=0}^{\infty}\frac{1}{16^i}\left(\frac{1}{(8i+1)^2} + \frac{1}{2}\frac{1}{(8i+2)^2} + \frac{1}{8}\frac{1}{(8i+3)^2}\right.\\ &\qquad\left. - \frac{1}{64}\frac{1}{(8i+5)^2} - \frac{1}{128}\frac{1}{(8i+6)^2} - \frac{1}{512}\frac{1}{(8i+7)^2}\right). \end{aligned}$$

Thus, digits of both G and π may be computed in the same fashion, and we might hope that the formula sheds some light on the normality of Catalan's constant. [Recall that a number is 'normal' if its digits occur with equal frequency.]

8. SIN, REDEMPTION, AND CAUTIONARY TALES

> The object of mathematical rigor is to sanction and legitimize the conquests of intuition, and there was never any other object for it.
>
> —J. Hadamard, quoted in [**24**]

Experimental mathematics cannot supplant rigorous mathematics. Dropping the latter for the former would indeed be a 'sin'. We have seen at least one example of

a false computer generated conjecture—namely the egyptian fractions example in Section 1—and we could come up with many more [5]. Experimental mathematics is, however, a good supplement to rigorous mathematics. It can enrich our subject and, when used with discipline, can significantly assist mathematical discovery. We have also seen examples where the computer can assist with the proof.

As a final demonstration, consider the power series

$$J(x) = \sum_{n_1 > n_2 > 0} \frac{x^{n_1}}{n_1^2\, n_2}.$$

In [8], a functional relation was sought in pursuit of a proof of the identity $J(1) = 8J(-1)$. For $0 \le x \le 1$,

$$J(x) = \int_0^x \frac{\ln^2(1-t)}{2t}\,dt = \zeta(3) + \frac{1}{2}\ln^2(1-x)\ln(x) + \ln(1-x)\operatorname{polylog}(2, 1-x) - \operatorname{polylog}(3, 1-x).$$

It can be shown that

$$J(-x) = -J(x) + \frac{1}{4}J(x^2) + J\left(\frac{2x}{x+1}\right) - \frac{1}{8}J\left(\frac{4x}{(x+1)^2}\right). \tag{22}$$

This relation was found, once the ingredients were determined by inspection, by evaluating (22) (actually, a version of it with undetermined coefficients) at a random point and then using LLL. Another successful strategy is to evaluate each J function at enough specific values of x to enable one to solve linear equations for the unknown coefficients.

If $L(x)$ and $R(x)$ denote the left-hand and the right-hand sides of (22), respectively, then computer manipulations (under the assumption $0 < x < 1$) show that $dL/dx = dR/dx$: mechanically differentiating both sides and using `simplify` reduces the difference between the two to zero. Observing that $L(0) = R(0) = 0$ completes a proof of (22).

8.1 Knowing 'the answer' might limit us. We are all familiar with examples of the value of 'doing things ourselves'. It is now trivial in most computer algebra systems (CAS) to compute very large values of the *partition function* with little or no thought, directly from the generating function

$$P(q) = \frac{1}{\prod_{n=1}^{\infty}(1-q^n)}.$$

The well-known exact finite series for values of the partition function, due to Radamacher [25], and its wonderful infinite, asymptotic precursor due to Ramanujan and Hardy, might well have seemed less worthy of discovery, had CAS been available then. We must be careful to ensure that our use of new tools neither limits us to what they can find for us nor supresses our interest in things easily computed.

This really will require attention: for example, the authors of [9] report in their conclusions that had they been aware of the answers in the Encyclopedia, they might not have bothered to prove what they did—and their results went beyond those in the Encyclopedia!

ACKNOWLEDGMENTS. This work was supported by NSERC. David Boyd, Mark Giesbrecht, David Jeffrey, and Glen Ord were all helpful, but the contributions of Petr Lisoněk were exceptional.

APPENDIX A. THE LAMBERT W FUNCTION IN BRIEF. If you have used MAPLE to solve transcendental equations, you may already have encountered the Lambert W function, defined by (3). The history and some of the properties of this remarkable function are described in **[14]**. This function provides a beautiful new look at much of undergraduate mathematics, in addition to some new puzzles of intrinsic interest.

Here are some of the elementary properties of W.

1. On $0 \le x < \infty$ there is one real-valued branch $W(x) \ge 0$ (see Figure 1). On $-1/e < x < 0$ there are two real-valued branches. We call the branch that has $W(0) = 0$ the principal branch. On this branch, it is easy to see that $W(e) = W(1 \cdot e^1) = 1$.
2. The derivative of W can be found by implicit differentiation to be

$$\frac{d}{dx}W(x) = \frac{1}{(1 + W(x))e^{W(x)}} = \frac{W(x)}{(1 + W(x))x}$$

where the second formula follows on using $\exp W(x) = x/W(x)$, and holds if $x \neq 0$. We may use the first formula to find the value of the derivative at $x = 0$, and we see the singularity is just a removable one.

3. The function $y = W(\exp z)$ satisfies

$$y + \log y = z.$$

This function appears, for example, in convex optimization. Consider the *convex conjugate*, $f^*(s) = \sup_r rs - f(r)$, of the function $f(r) = r\ln(r/(1 - r)) - r$. Calculation shows that $f^*(s)$ is just $W(\exp s)$.

4. $W(x)$ has a Taylor series about $x = 0$ with rational coefficients. Similarly, $W(\exp z)$ has a Taylor series with rational coefficients about $z = 1$. MAPLE computes the first few terms to be

$$\begin{aligned} W(e^z) := 1 &+ \frac{1}{2}(z-1) + \frac{1}{16}(z-1)^2 - \frac{1}{192}(z-1)^3 - \frac{1}{3072}(z-1)^4 \\ &+ \frac{13}{61440}(z-1)^5 - \frac{47}{1474560}(z-1)^6 - \frac{73}{41287680}(z-1)^7 \\ &+ \frac{2447}{1321205760}(z-1)^8 - \frac{16811}{47563407360}(z-1)^9 \\ &- \frac{15551}{1902536294400}(z-1)^{10} + O\left((z-1)^{11}\right) \end{aligned}$$

Here is an exact formula for the coefficients of the nth derivative of $W(\exp z)$, containing second-order Eulerian numbers $\left\langle\!\left\langle {n \atop k} \right\rangle\!\right\rangle$ **[18]**. This formula comes from the following expression for the nth derivative of $W(\exp z)$, which is stated in **[14]**. Once the answer is known, the proof is an easy induction, which we leave for the reader.

The derivatives of $W(\exp z)$ are

$$\frac{d^n}{dz^n}W(e^z) = \frac{q_n(W(e^z))}{(1 + W(e^z))^{2n-1}}, \tag{23}$$

where $q_n(w)$ is a polynomial of degree n satisfying the recurrence relation

$$q_{n+1}(w) = -(2n-1)wq_n(w) + (w + w^2)q_n'(w), \quad n > 1 \tag{24}$$

and having the explicit expression

$$q_n(w) = \sum_{k=0}^{n-1} \left\langle\!\!\left\langle {n-1 \atop k} \right\rangle\!\!\right\rangle (-1)^k \, w^{k+1}. \tag{25}$$

If $n = 1$ we have $q_1(w) = w$, and it is convenient to put $q_0(w) = w/(1+w)$; this isn't a polynomial, but it makes things work out right. This means that our series for $W(\exp z)$ about $z = 1$ is just

$$W(e^z) = \sum_{n \geq 0} \frac{q_n(1)}{n! 2^{2n-1}} (z-1)^n. \tag{26}$$

REFERENCES

1. G. Almkvist and A. Granville, Borwein and Bradley's Apéry-like formulae for $\zeta(4n+3)$, *Experiment. Math.* 8 (1999) 197–204.
2. David Bailey, Peter Borwein, and Simon Plouffe, On the rapid computation of various polylogarithmic constants, *Math. Comp.*, 66 (1997) 903–913.
3. Jonathan M. Borwein and Peter B. Borwein, *Pi and the AGM*, John Wiley & Sons, New York, 1987.
4. Jonathan M. Borwein and Peter B. Borwein, A cubic counterpart of Jacobi's identity and the AGM, *Trans. Amer. Math. Soc.* 323 (1991) 691–701.
5. Jonathan M. Borwein and Peter B. Borwein, Strange series evaluations and high precision fraud, *Amer. Math. Monthly* 99 (1992) 622–640.
6. Jonathan M. Borwein, Peter B. Borwein, and Frank G. Garvan, Some cubic identities of Ramanujan, *Trans. Amer. Math. Soc.* 343 (1994) 35–47.
7. Jonathan M. Borwein and David M. Bradley, Empirically determined Apéry-like formulae for zeta($4n+3$)}, *Experiment. Math.* (1997) 181–194.
8. Jonathan M. Borwein, David M. Bradley, David J. Broadhurst, and P. Lisoněk, Special values of multidimensional polylogarithms, *Trans. Amer. Math. Soc.* (to appear).
9. Jonathan M. Borwein and Kwok-Kwong Stephen Choi, On the representations of $xy + xz + yx$, Technical Report 98-119, http://www.cecm.sfu.ca/preprints, 1998.
10. Jonathan M. Borwein and Petr Lisoněk, Applications of integer relation algorithms, *Discrete Math.* (Special Issue for FPSAC 1997), to appear.
11. Peter Borwein and Loki Jörgensen, Visible structures in number theory, *http://www.cecm.sfu.ca/~loki/Papers/Numbers/*, 1998.
12. David J. Broadhurst, Polylogarithmic ladders, hypergeometric series and the ten millionth digits of $\zeta(3)$ and $\zeta(5)$, *preprint*, January 1998.
13. Robert M. Corless, Error backward, in *Proceedings of Chaotic Numerics*, volume 172 of *AMS Contemporary Mathematics*, 1994, 31–62.
14. Robert M. Corless, Gaston H. Gonnet, David E. G. Hare, David J. Jeffrey, and Donald E. Knuth, On the Lambert W function, *Adv. Comput. Math.* 5 (1996) 329–359.
15. Robert M. Corless and David J. Jeffrey, graphing elementary Riemann surfaces, *ACM Sigsam Bulletin: Communications in Computer Algebra* 32(1) (1998) 11–17.
16. Robert M. Corless, David J. Jeffrey, and Donald E. Knuth, A sequence of series for the Lambert W function, in W. Küchlin, editor, *Proceedings of ISSAC '97, Maui*, 1997, pp. 197–204.
17. H. R. P. Ferguson, D. H. Bailey, and S. Arno, Analysis of PSLQ, an integer relation finding algorithm, Technical Report NAS-96-005, NASA Ames Research Center, Moffett Field, CA, April 1996.
18. Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete Mathematics*, Addison-Wesley, 1994.
19. D. F. Griffiths and J. M. Sanz-Serna, On the scope of the method of modified equations, *SIAM J. Sci. Stat. Comput.* 7 (1986) 994–1008.
20. G. H. Hardy, *Ramanujan*, Chelsea, New York, 1940.
21. Gilbert Labelle, Sur l'inversion et l'itération continue des séries formelles, *Eur. J. Combinatorics* 1 (1980) 113–138.
22. A. K. Lenstra, H. W. Lenstra Jr., and L. Lovász, Factoring polynomials with rational coefficients, *Math. Ann.*, 261 (1982) 515–534.

23. George Marsaglia and John C. Marsaglia, A new derivation of Stirling's approximation to $n!$, *Amer. Math. Monthly* 97 (1990) 826–829.
24. George Polya, *Mathematical discovery: On understanding, learning, and teaching problem solving*, John Wiley & Sons, 1981.
25. H. Radamacher, On the partition function $p(n)$, *Proc. London Math. Soc.*, 43 (1937) 241–254.
26. Ed Regis, *Who Got Einstein's Office?*, Addison-Wesley, 1986.
27. Daniel Richardson, Some unsolvable problems involving elementary functions of a real variable, *J. Symbolic Logic* 33 (1968) 511–520.
28. N. J. A. Sloane and Simon Plouffe, *The Encyclopedia of Integer Sequences*, Academic Press, 1995.
29. Michal Trott, Visualization of Riemann surfaces of algebraic functions, *Mathematica in Education and Research* 6(4) (1997) 15–36.

JONATHAN M. BORWEIN is the director of the Centre for Experimental and Constructive Mathematics at Simon Fraser University in Vancouver, BC. He was an Ontario Rhodes Scholar (1971), D.Phil (Oxford) 1974 under Michael Dempster, the 1987 Coxeter-James lecturer, won the 1988 Atlantic Provinces Gold Medal for Sciences, and the Shrum Professorship of Science at Simon Fraser University. His research background includes (nonsmooth) optimization theory, functional and classical analysis, computational complexity, and, of course, π; however, lately he has been seduced by the transformation of mathematics by computation and the internet. He also enjoys good food, good whisky, good books, and good fast conversation.
Centre for Experimental and Constructive Mathematics, Simon Fraser University, Vancouver, B.C., Canada, V5A 1S6
jborwein@cecm.sfu.ca

ROBERT M. CORLESS, Ph.D. UBC 1987 (Mech. Eng.), is Chair of the ACM Special Interest Group on Symbolic and Algebraic Manipulation, Chair of the ISAAC Steering Committee, Deputy Director of the Ontario Research Centre for Computer Algebra, and Professor in the Department of Applied Mathematics at The University of Western Ontario. His research interests include computer algebra (especially symbolic-numeric algorithms for polynomials), numerical analysis, dynamical systems, and flow-induced vibrations. He is a member of SIAM, CMS, a life member of The Canadian Applied and Industrial Mathematics Society, and has been an MAA member since he won a one-year membership in his final undergraduate year in 1980. Like Jon, he enjoys good food, wine, books, music, and conversation.
Department of Applied Mathematics, University of Western Ontario, London, Ontario, Canada, N6A 5B7
Rob.Corless@uwo.ca

Reform, Tradition, and Synthesis

Thomas W. Tucker

The recent debate in the mathematical community about calculus instruction is not the first struggle between reform and tradition, and it won't be the last. Perhaps the two sides in this case may be much closer to agreement than the rhetoric indicates. My goal is to make a few remarks about some of the issues that have divided the two sides: technology, lecturing, drill, rigor, algebra, choice, and outcomes. For most of these issues, there are stances usually attributed to the traditionalists and reformers. For example, reformers may favor collaborative learning, while traditionalists prefer lectures. I don't think such contrasts are necessarily accurate, and I hope my remarks might initiate a dialogue to reach some middle position, a synthesis of tradition and reform. I should acknowledge that, although I have spent most of the last twelve years in the reform camp (through work in the MAA and in the Calculus Consortium based at Harvard), I am a timid reformer and I make no claims that my views represent anything but my own opinion.

Before I proceed I offer a brief apology for the use of the word "reform," whose connotations are not nice (e.g., "reform school"). I was always a little hesitant to use the word in the early days of calculus reform after the 1986 Tulane Conference, which itself avoided the word as much as possible. Unfortunately, no better term came along and we are stuck with it now. I thought of writing this whole article spelling the word "re-form" instead of "reform," but that seemed too precious. So it's "reform," warts and all.

Technology. I asked my multivariable calculus class yesterday if they knew the sines or cosines of any special angles, whether the numbers 2 or root 3 sounded familiar. Only about a third of the class raised a hand. I know this would not have happened in a calculus class thirty years ago. I mourn the loss of this lovely bit of knowledge. The widespread use of graphing calculators is to blame, of course, and it can get worse. As graphing calculators with symbolic manipulation become more widely used, I share the fears of many mathematicians that a question about the derivative of sine or cosine could draw equally blank looks from my class. If it's in the machine, why memorize it? I sympathize with mathematicians who ban graphing calculators in their classes. When I was on the AP Calculus committee a few years ago, I strongly supported having sections of the test where graphing calculators are not allowed. On the other hand, I also strongly supported allowing them, or even requiring them, on other parts of the test.

Here is why. It pays to heed history: Technology always wins. The world may have been better when people walked instead of driving cars, but that is irrelevant. As long as there is gas, people will drive cars, and what I really care about is that they drive them sensibly. The mathematical world may have been better when people did arithmetic or graphed functions on paper or in their head instead of on a calculator, but that is irrelevant. As long as there are batteries, students will use calculators, and what I really care about is that they use them sensibly. So I

allow them in my classes and have learned to appreciate my students' facility and inventiveness. When my students misuse their calculators or something unexpected happens, I have an opportunity to give them some important advice or talk about an interesting mathematical phenomenon. Pretending something doesn't exist is not a good teaching strategy. For many of my students, graphing calculators are as much a part of their intellectual constitution as pencil and paper, and I have to learn to deal with it.

Computers seem to be less an issue. Indeed, I suspect there are some calculus courses that require computer laboratories or assignments but ban graphing calculators. This is actually quite understandable. Most mathematicians spend enough time around computers, in their everyday life or even in their research, that it seems natural to use computers in their teaching as well. On the other hand, most college or university mathematicians have spent no time at all with a graphing calculator and are not inclined to spend the start-up time of an hour or two to learn, especially since they are unlikely to use graphing calculators on a regular basis outside the classroom. A bridge is needed for this gap between mathematics students (and secondary school teachers) on the one side and college faculty on the other. Indeed, I think the role of hand-held devices in mathematics education, from college right down to kindergarten, needs to be studied and discussed far more than at present. For example, is long division with pencil and paper still a necessary skill? No one seems to be willing to entertain the notion that it is not, and until someone does, I don't think there will be an honest discussion.

Lecturing. Let's cut to the chase. Do I lecture? Yes. All the time? Just about. Do I believe that students learn by talking to each other? Absolutely, because I myself learn best by talking with other mathematicians, even when we have little idea what we are talking about. My implementation of collaborative learning is low-key. Once or twice a week, I have a pair of students present a homework problem on the board (they know ahead of time who their partner is and which problem they have to do). In a class of 35, this gets everyone to the board at least once during the semester at the cost of 10 minutes a week. This gets control of the blackboard out of my hands for a few minutes and forces two students to talk to each other outside of class. I also distribute a class list, which includes email addresses, phone numbers, and dorm rooms, so everyone can find someone to hook up with. I never make available a solutions manual so students are forced to talk to someone else when they are confused. The result is that usually a little more than half the students in my classes work on their homework in groups of two or more. I wish it were more and I harangue them as much as I can, but loners may be happier as loners and I can't change that. All I know is that I was a loner myself in my undergraduate courses and couldn't have been more unhappy (mathematically). Sometimes all it takes is a nudge in the right direction.

Classroom formats with little lecturing can be wonderful, but the evolutionary forces that brought us the lecture format havn't gone away. Lectures are here to stay. The real issue is how to get students talking with each other, and there are lots of mechanisms for doing that.

Drill. A colleague of mine has said "There are some things you should do with your spine rather than your brain." I agree. Students should be able to take derivatives of most elementary functions without having to think about it, with their spine. Again, I am delighted that the AP Calculus exam has a multiple choice section where calculators are not allowed and "spinal" manipulations can be tested without interference from calculators that can take derivatives symbolically. To do

this, students need drill. The question is determining when you have reached the point of diminishing returns. If I drill my students on differentiation all semester, there will still be some who make mistakes on a four-deep chain rule. In the meantime, think of the other things I could have done.

Another colleague has said "Better rote learning than no learning." I used to agree, mostly because I think memorization is good for the mind. I am not so sure, however, whether this is true in mathematics. The belief that mathematics is just formulas, a belief that studies show American students hold and Japanese students do not, undermines everything mathematics educators are trying to do. Some rote, some drill, fine, but it better be less than half of what is taught and tested, or else it isn't mathematics anymore.

Rigor. When it comes to theory in calculus courses, mathematicians surrendered a long time ago. There is almost no theoretical content at all in the compendium of calculus final exams given in the 1987 MAA Notes Volume, *Calculus for a New Century*. Despite the talk that one can learn mathematics (or any science for that matter) only by doing it, when it comes to theory, students have no hands-on activity. Students may see correct definitions and proofs but they don't do them. I understand why the debate over rigor in calculus instruction has been so bitter: mathematicians have conceded so much since the heights of abstraction reached in the new math era of the 1960's, that they cling to what little formalism remains. I hope instead there is a serious effort to reclaim the high ground.

I think calculus students should do proofs. The word "prove" should appear in problems. One should be careful, however, about what students are asked to prove. In mathematical research, proof is a tool used to answer questions where the issue is in doubt. Asking for an epsilon-delta proof that a certain limit is what we know it must be is guaranteed to irritate and confuse students. Ask instead for proofs in situations where there is doubt. For example: Prove or disprove that if two functions are both concave up on an interval, their sum is concave as well. I know a few other examples (a couple have appeared on AP exams), but many more are needed.

I also think students should write sentences and paragraphs in which they use formal mathematical terminology correctly. The mathematical content does not have to be deep; a full discussion of the graphical behavior of some function is enough. The culture shock that hits mathematics majors in their first theory course is not just the abstraction. It is that arguments are to be written in logically coherent sentences and paragraphs, not strings of equations as usually is the case in a calculus class. At the very least, students should be asked frequently to explain what they think they are doing. Although some reform projects have worked very hard on improving student writing, I hardly think of this as a reform issue. Students need to write.

Algebra. I guess this is the one area where I am most fervently a reformist. Algebra is one of the most powerful intellectual tools known to mankind. Computers could not operate without algebraic representations of functions. Students can, however, get the impression from calculus (and earlier mathematics courses) that algebra and mathematics are synonymous. That is not good. I have already noted how American students seem to think mathematics is just formulas. Far worse, if mathematics is algebra, then it must be irrelevant to most students' lives. Just read the *New York Times* for a month, every page, and tell me how often you encounter an algebraic equation or formula. There is plenty of mathematics there in numbers, tables, graphs, or verbal descriptions, but nary an x or y in sight. I often

think that my own algebraic manipulative skills stay honed only because I teach calculus; I certainly don't use those skills much in my research.

Functions in a calculus course should be represented by tables of values, graphs, and verbal descriptions, as well as algebraic formulas. This does not water down the course. Non-algebraic reasoning and communication is not "softer" than algebraic, any more than geometry is softer than algebra. Interestingly enough, the inclusion of non-algebraic viewpoints seems to be one aspect of calculus reform that has gained acceptance. It is often the way new editions of many traditional texts most resemble reform texts, and it has also become part of the guidelines for the construction of many standardized mathematics tests.

Choice. Back around 1990, Peter Lax proposed to the American Mathematical Society the following resolution that might act like a stick of dynamite to break up the logjam in curricular diversity: "Requiring a professor to teach from a common textbook or for a common exam is an abridgment of academic freedom." I remember this sounded awfully revolutionary. I believe Peter was careful to say "professor," and there may have been some weasel words, like "qualified" or "tenured" professor, but still it seemed common sense that some sort of uniformity is needed in a multiple section calculus course taught by professors, post-docs, adjuncts, and graduate students. Nowadays, it is becoming more common to see fewer common exams and even different textbooks in different sections of a calculus course. This is a reasonable compromise when departments (such as my own) cannot reach a consensus on how to teach calculus.

I am still not sure how I feel about this. Diversity is better, I know. Even the most traditional calculus instructor has bemoaned at least once the lack of variety in textbooks. For a few years in the early days of calculus reform, there really was some choice; now there is still some diversity, although less than before as the more radical texts are remaindered by publishers. On the other hand, students are prone to making invidious comparisons, and it is a lot easier for everyone if all sections of a multisection course look the same. Also, making up and grading common exams is a source of departmental camaradie; many reform efforts focus on the social aspects of teaching and learning, and common syllabi and exams build community, both among faculty and among students. In general, it is probably better for a department to reach some compromise consensus for its calculus courses. Allowing each instructor to go his or her own way, with only an agreement over the core content, should be a last resort.

Outcomes. Reform courses have been under pressure to assess their success. You can't say something is better without backing it up with data. I have always viewed this as a red herring. First, traditional courses do almost no assessing of outcomes other than student performance on the final exams; I doubt that the pass/fail rate on a final exam is viewed as a reasonable form of assessment. Second, most reformers end up working 16 hour days to prepare new materials (the usual criticism of reform courses is that they are way too labor intensive) and have little time for extensive assessment. Finally, most of the comparisons I know between traditional and reform courses at the same institution are not controlled experiments: even when the students are assigned randomly to different sections, the instructors are not. When reform courses come out looking better on common exams, perhaps it is because the instructors who choose to teach the reform versions are not typical instructors.

Nevertheless, the call for assessments is useful. It is a good idea to think hard about what students take with them from a course, in terms of not only content but

also experience. For example, one should question the choice of content for a first semester calculus course that does not include the exponential and natural log functions; after all, the course is probably terminal for half the class. In terms of experience, one should ask questions of a calculus course that could be asked of any course: Did students have to write? Did they speak to an audience? Did they have an opportunity to work on some significant project independently? Did they acquire a viewpoint or skills that are applicable in a wide variety of circumstances? Did they work with others? Did they have to find and evaluate information for themselves, from a library or the web? The outcomes of a calculus course should be viewed in the context of the entire college curriculum.

Community. It has been observed that one thing reform has accomplished in the last ten years is the creation of a community of mathematicians who share a common interest in mathematics education. The more people who feel they are part of this community, the better. That is why it is so important for both reformers and traditionalists to see their common interest: they both want their students to learn and appreciate mathematics.

THOMAS W. TUCKER received his BA from Harvard University in 1967 and Ph.D. from Dartmouth in 1971. After two years teaching at Princeton University, he came to Colgate University in 1973, where he is now the Charles Hetherington Professor of Mathematics. He has been active in calculus reform as chair from 1988 to 1992 of the MAA committee on Calculus Reform and the First Two Years (CRAFTY) and as a member of the NSF-supported Calculus Consortium based at Harvard. His research interests are in low-dimensional topology and topological graph theory.
Colgate University, Hamilton, NY 13346
ttucker@mail.colgate.edu

Editor's Note: The preceding article by Thomas Tucker and the following article by Steven Krantz were solicited to present a collegial discourse about calculus reform. Each author was encouraged to comment on positive attributes of the 'other side' and to be honest about problems on 'their side'.

You Don't Need a Weatherman to Know Which Way the Wind Blows

Steven G. Krantz

I am moderately well-known as a complex analyst, but I seem to be almost pathologically well-known as an avatar of traditionalist teaching. The latter attribute stems no doubt from my having penned the book *How to Teach Mathematics*. The discussions pursuant to the appearance of that book have caused all of us to rethink our positions. Certainly my ideas have evolved. Have no fear: I still value traditional methods of teaching. But I have come to appreciate many of the reform ideas as well. Nobody wants to be told that the tried-and-true methods that he or she has been using for several decades are no longer valid. But any well-educated person who is capable of critical thinking surely knows that a skill worth learning is also one that is worth rethinking and refining and developing. What do the reformers have to offer that might appeal to such an individual?

Perhaps the most compelling, yet disturbing, assertion that I have heard from the reformers is this: "It's not just that lecturing doesn't work with today's students. In fact lecturing has never worked." Can this be true? Sadly, you and I are ill-equipped to judge. As professional mathematics instructors, we are the survivors in a rather arcane evolutionary process. We were always good at learning—particularly at learning mathematics. Our mathematical and scholarly abilities raised us to such a level that we were relatively immune to what teaching methods were being used, or what personality quirks the teacher had, or what medieval textbook was being foisted upon us. Alas, most students don't fit that mold. It is valid, and appropriate, to pose the question of whether there are teaching techniques that are more effective than lecturing *in teaching an average student of average ability*.

I still lecture; on days of extraordinary hubris, I think I'm pretty good at it. But I endeavor to create the illusion in my classroom that the students and I are actually carrying on a dialogue, that we are developing the ideas together. In my own way, I am enabling my students to engage in group work, and to participate in discovery learning. I may not be a card-carrying reformer, but I have been influenced by the reform tenets.

In the past few years I have become convinced that lower division mathematics should be a laboratory science. Chemists and biologists have known for lo these many years that labs are an effective way to make ideas concrete for the student. They are a way to enable discovery learning. Why has mathematics remained out of the loop?

One obvious reason is that accessible and affordable high speed digital computing has been unavailable until fairly recently. Quality software—that is of interest to the mathematician—did not exist. But things have changed: most math departments are full of computer equipment and also full of exciting new software tools such as `Mathematica` and `Derive` and `Axiom` and `Maple`. Do you find it difficult to explain to your students why the method of Lagrange multipliers works? Or why the gradient of a function of three variables is always orthogonal to the level sets? Or why Simpson's rule converges more rapidly than the trapezoid

rule? Couldn't well-constructed computer labs bridge this gap, and help students of average ability to understand why and how mathematics works?

In the past I have been guilty of asking:

- How can students discover mathematical facts if they have no knowledge base and no technical training?
- How can students work in groups when nobody in the group knows what he or she is talking about?
- How can students formulate conjectures if they don't know anything?

These questions are not entirely off-base. But they are a bit cranky. And well-thought-out laboratories may provide at least a partial answer to all of them. A student might discover a mathematical fact if a lab activity is designed to lead him or her to it. Students might discuss and collaborate profitably if (computer-aided) material is put before them that will stimulate such interaction. A highly trained person—say a Ph.D. in mathematics—needs very little grist, and almost no catalyst, to get his mill grinding. A young student needs considerably more, and interaction with the computer can help. It is difficult for a person lacking a highly developed intellectual framework to formulate conjectures; but a good computer lab can help the student to build a short-term framework that will lead to interesting queries.

A good teacher does three things for his/her students:

(1) Sets a pace for the students;
(2) Teaches the students to read;
(3) Engages the students in the learning process.

It is item **(3)** that causes most of us the greatest frustration and discomfort. Why won't our students talk to us? Why don't they show any interest? Why is class attendance so poor? Why is there no sense of curiosity or excitement in the typical calculus classroom?

I'm sorry to say it—I know that nobody wants to hear it—but lectures, in and of themselves, are not by nature engaging or exciting. At least not for eighteen-year-olds. This has been one of the chief messages of the reform movement and, in essence, I think that the message is correct. I have learned to use my own lectures as an effective tool. I fill the room with myself; I get my students to talk to me. Under my guidance, the students shout out conjectures, and they *help me to construct the lesson*. This is a skill that I have honed over more than one quarter of a century of teaching. But it is a great deal of work to develop such a skill. Not all of us are born with such skill or such dedication, nor do we all have the inclination to learn it. A reasonable alternative is to say, "Lectures are not working; let's try something else."

I don't buy in to *that* particular conclusion. I have learned to make my lectures work for me. And they work for my students too. But each mathematics instructor must find his or her own means of getting students involved in the learning process, of helping them to become educated. The reformers have put before us a menu of possibilities—including group work, discovery learning, computer labs, and other techniques too—that are well worth exploring. Take those that appeal, sample some others. Keep the ones that work. And then move on.

One of the more controversial tenets of reform is that we should reduce the role of drill in our classrooms, that we should soft-pedal rigor and theory, and that we should instead concentrate on *concepts*. [Certainly you cannot claim to the world that you have written a reform calculus book unless the word "concepts" appears

in your title.] How is a died-in-the-wool traditionalist to come to terms with these notions?

I am convinced that our freshmen are very bright, but they do not have the intellectual equipment to appreciate a genuine mathematical proof. Those who have taken a high school course in Euclidean geometry in the past ten years did *not* have the course that some of us experienced thirty years ago. Modern high school geometry texts minimize proofs (and stress concepts!). You and I have intensive training in the discourse of mathematics. When I write a proof—that I want you to read—then I prepare it in the accepted form that I have been trained to produce, so that you will both appreciate it and believe it. Our freshmen are *not* privy to this discourse.

In a class full of freshmen, I find it appropriate to say "Here is a picture that illustrates why this is true" (when I am explaining, for example, the Fundamental Theorem of Calculus) or "Here is an example that shows why this works" (when I am explaining why $\det(A \cdot B) = (\det A) \cdot (\det B)$ or "Here is an analogy that will help you to believe this formula" (when I am explaining the Chain Rule). You have to speak to people in their own language. For freshmen that language is English. If the math curriculum is well-constructed, then by the time that the student is a junior he or she will have learned *mathematical argot*; at that time we can present such a student with a proof, and he/she will appreciate it (and believe it). Prior to that, we should resist.

Do I teach concepts? Who wouldn't? On the one hand, we teach students *technique*. For instance, when we teach maximum/minimum problems we show them how to actually *do* such problems; on the other hand there is a concept (due to Fermat) behind the technique, and we teach that as well. Concepts without technique are hollow. Technique without drill is meaningless. Most reformers that I know would agree. There is some debate over whether drill or concepts should come first. I leave that to the individual: there are many worthy and productive paths that lead to the same goal.

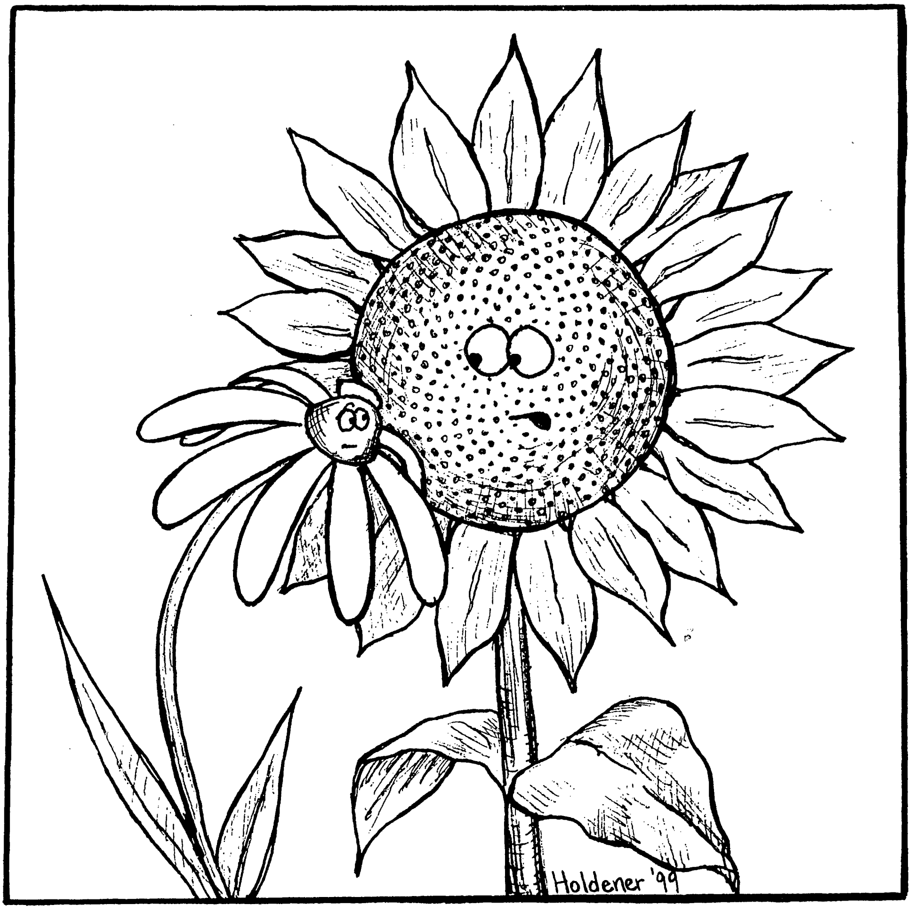
For many years we have all known, in the backs of our minds, that our students cannot write. They hand in homework assignments that bear scant resemblance to anything more than incoherent gibberish. The reformers—especially the Harvard group—have helped us to realize that writing has a deserved place in the mathematics classroom. And I'm talking about real writing here, with sentences and paragraphs and overall organization. The accident at Three Mile Island occurred in large part because the engineers at that power plant could not communicate their concerns to the governor of Pennsylvania. I wonder how many of those engineers were our calculus students?

Good writing and clear thinking are inexorably linked. Certainly we all want our students to be clear thinkers. One sure way to help them develop in that direction is to teach them to write, to organize their thoughts, to judge their audience, to argue a point. It is just a bit too facile for us to object that all these reform techniques take more time and more effort on the parts of the instructors. Of course they do. Anything worthwhile requires a great deal of effort. Once we have decided that these methodologies are worthwhile, and worth trying, then we can find practical methods for implementing them.

Reform always works in the hands of the reformers. For everyone else, reform is an object lesson and a crucible for experimentation. We will all be better off when we realize that reformers and traditionalists are after the same grail: to enable our students to appreciate and to learn and in the end perhaps to love mathematics. We want to give them the grounding they need in mathematical techniques and

concepts so that they can go on to advanced study in any area they might choose to pursue, whether it be engineering or epidemiology or even mathematics. We want them, as part of their education in Western thought, to understand the mathematical method. These realizations should make it easy for reformers and traditionalists to work together. Let us find the means to do so.

STEVEN G. KRANTZ received his B.A. degree from the University of California at Santa Cruz in 1971. He received the Ph.D. from Princeton University in 1974. Krantz has taught at UCLA, Princeton University, Penn State University, and Washington University in St. Louis. He has been a visiting professor at Princeton University, the University of Paris, the University Paul Sabatier, the University of Umea, Uppsala University, the University Autonoma de Madrid, the Mathematical Sciences Research Institute, the Institute for Advanced Study, and Beijing University. Krantz has received the UCLA Alumni Association Distinguished Teaching Award, the Kemper Prize, the Chauvenet Prize of the MAA, and the Beckenbach Book Award of the MAA.
Washington University in St. Louis, Campus Box 1146, One Brookings Drive, St. Louis, Missouri 63130
sk@math.wustl.edu



"Ever notice that the number of legs on an animal is always a number from the sequence {0, 2, 4, 6, 8,...}?"

Contributed by Judy Holdener, Kenyon College

The Weyr Characteristic

Helene Shapiro

1. INTRODUCTION. The Jordan canonical form is a well-known and standard topic in linear algebra. It is thoroughly covered in many texts on linear algebra and abstract algebra. The purpose of this article is to publicize a different approach to the canonical form problem introduced by Eduard Weyr in 1885 [**28**], [**29**]. Several older books ([**15**, pp. 73–74] and [**16**, pp. 117–118]) mention Weyr characteristics but it does not appear in recent linear algebra texts. The basic idea of Weyr's approach is useful in several areas, such as describing algorithms for computing the Jordan form in a stable manner ([**8**], [**13**], and [**18**]), and in developing canonical forms for matrices under unitary similarity ([**2**], [**14**], [**21**], and [**22**]), but Weyr's papers are rarely referenced and the sequence of numbers we call the Weyr characteristic is not named. Thus, while Weyr's work seems to be little known, his basic idea has been rediscovered and used several times. I first learned of the Weyr characteristic from Hans Schneider, when I was a post-doc at the University of Wisconsin in 1980. Schneider and others have studied the relationship between the Weyr characteristic and the singular graph of an M-matrix ([**9**], [**10**], [**17**], and [**19**]).

In this paper we define the Weyr characteristic and discuss its connection with the so-called "staircase" forms used in numerical linear algebra to determine the Jordan form in a stable manner. There is a simple relationship between the Weyr characteristic and the better known Segre characteristic, which is associated with the Jordan canonical form. This relationship leads to a quick derivation of Weyr's canonical form from the Jordan canonical form; we also present a proof that is independent of the Jordan canonical form, as Weyr did in his original paper.

The Jordan canonical form gives a canonical form for square matrices under the equivalence relation of similarity. It can be used whenever the field contains the eigenvalues of the matrix; typically, one is interested in matrices over the field of complex numbers. The Jordan canonical form of a square matrix A is a direct sum of square submatrices, called *Jordan blocks*. Each such block has an eigenvalue of A in the diagonal entries, a line of 1's along the superdiagonal, and zeros in all other entries, as shown in Figure 1.

$$\begin{pmatrix} \alpha & 1 & 0 & 0 & \cdots & 0 \\ 0 & \alpha & 1 & 0 & \cdots & 0 \\ 0 & 0 & \alpha & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \alpha & 1 \\ 0 & 0 & 0 & 0 & \cdots & \alpha \end{pmatrix}$$

Figure 1. A Jordan block with eigenvalue α.

There is at least one Jordan block for each eigenvalue of A and there may be several Jordan blocks for a single eigenvalue. The list of the non-increasingly

ordered sizes of the blocks belonging to a given eigenvalue α is called the *Segre characteristic* of A relative to α. The Jordan canonical form displays all the information needed to know the algebraic structure of a linear transformation. The eigenvalues appear on the main diagonal, and the Segre characteristic reflects the action of the transformation on the generalized eigenspaces. To quote Golub and Wilkinson [**8**, p. 5768], "From the standpoint of classical algebra, the algebraic eigenvalue problem has been completely solved. The problem is the subject of classical similarity theory, and the fundamental result is embodied in the Jordan canonical form."

Weyr's canonical form is a block triangular matrix in which the diagonal blocks are scalar matrices (that is, scalar multiples of identity matrices), the superdiagonal blocks contain identity matrices augmented by rows of zeros, and all the other blocks are zero. The list of the non-increasingly ordered sizes of the diagonal blocks corresponding to an eigenvalue α is called the *Weyr characteristic* of A relative to α. These numbers are determined by the dimensions of the nullspaces of powers of $(A - \alpha I)$; we give precise definitions later. For example, if the Weyr characteristic of A corresponding to α is $(7, 5, 2, 2)$, then the block of the Weyr canonical form of A corresponding to α would have the form shown in Figure 2.

Weyr's approach is related to methods developed in numerical linear algebra for computing the complete eigenstructure of a matrix. While one can derive the Jordan canonical form using an algorithmic approach [**4**], there are numerical reasons to avoid direct computation of the Jordan form [**5**, p. 146]. In numerical computations, one must consider the effect of rounding errors and errors in the input. If the matrix is ill-conditioned with respect to the desired computation, or if the algorithm is not carefully designed, then small errors in the input or rounding errors may result in large errors in the output. Computing the inverse of a matrix that is close to being singular, or applying a similarity that is close to a singular matrix can lead to disaster. It is better to use algorithms that involve only orthogonal or unitary transformations. Algorithms developed by Kublanovskaya [**13**], Ruhe [**18**], and Golub and Wilkinson [**8**] for computing the Jordan canonical form of a matrix in an efficient and stable manner use unitary transformations to transform the matrix to a block triangular, or "staircase" form, in which the block sizes correspond to the Weyr characteristic. These algorithms are typically described in terms of QR factorizations, and/or singular value decompositions, but

$$\left(\begin{array}{ccccccc|ccccc|cc|cc}
\alpha &&&&&&& 1 &&&&&&&& \\
& \alpha &&&&&&& 1 &&&&&&& \\
&& \alpha &&&&&&& 1 &&&&&& \\
&&& \alpha &&&&&&& 1 &&&&& \\
&&&& \alpha &&&&&&& 1 &&&& \\
&&&&& \alpha && 0 & 0 & 0 & 0 & 0 &&&& \\
&&&&&& \alpha & 0 & 0 & 0 & 0 & 0 &&&& \\
\hline
&&&&&&& \alpha &&&&& 1 &&& \\
&&&&&&&& \alpha &&&&& 1 && \\
&&&&&&&&& \alpha &&& 0 & 0 && \\
&&&&&&&&&& \alpha && 0 & 0 && \\
&&&&&&&&&&& \alpha & 0 & 0 && \\
\hline
&&&&&&&&&&&& \alpha && 1 & \\
&&&&&&&&&&&&& \alpha && 1 \\
\hline
&&&&&&&&&&&&&& \alpha & \\
&&&&&&&&&&&&&&& \alpha
\end{array}\right)$$

Figure 2. The block of the Weyr canonical form corresponding to an eigenvalue α with Weyr characteristic $(7, 5, 2, 2)$.

in theoretical terms, these computations find the null spaces of powers of $(A - \alpha I)$, for each eigenvalue α. Related ideas also appear in Van Dooren's work ([**1**], [**25**], [**26**], and [**27**]) on computing the Kronecker normal form of a matrix pencil, $A + \lambda B$. We do not describe these methods here and refer the reader to the original sources for specific algorithms and an analysis of their stability and efficiency. Our aim is to present Weyr's basic theory and give some proofs that are motivated both by the methods used in the numerical algorithms and by Weyr's original presentation.

2. PRELIMINARIES. We work over an algebraically closed field F. The vector space $V = F^n$ is the space of column vectors of length n over F. If T is a linear operator on V, that is, a linear transformation from V to V, then T can be represented by an $n \times n$ matrix over F, relative to a choice of basis for V; the matrix representation depends on the choice of basis. If A and B are two $n \times n$ matrices that represent T, relative to two choices of basis, then A and B are related by the equation $B = P^{-1}AP$, where the nonsingular matrix P is the change of basis matrix. We say A and B are *similar*.

If F is the field of complex numbers C, we have the usual inner product on C^n. A square, complex matrix U is said to be unitary if $U^{-1} = U^*$ (the star denotes the conjugate transpose); this is equivalent to saying that the columns of U form an orthonormal basis for C^n with respect to the usual inner product. Applying a unitary similarity to A is equivalent to a unitary change of basis.

We frequently deal with matrices that are partitioned into submatrices that have special forms. If A is an $n \times n$ matrix, we may partition the rows of A into t sets consisting of the first n_1 rows, the next n_2 rows, and so on, finishing with the last n_t rows, where $n_1 + n_2 + \cdots + n_t = n$. Partitioning the columns of A in the same way breaks the matrix up into t^2 blocks, A_{ij}, where A_{ij} denotes the block formed from the ith set of rows and the jth set of columns. Note that A_{ij} is $n_i \times n_j$ and the diagonal blocks are square. If all blocks below the diagonal blocks are zero ($A_{ij} = 0$ for $i > j$) then we say A is *block (upper) triangular*. One can visualize the form of such a block triangular matrix as a staircase. If A_i denotes the ith diagonal block (A_{ii}) then we also say that A is $\mathcal{T}(A_1, A_2, \ldots, A_t)$ or write $A = \mathcal{T}(A_1, A_2, \ldots, A_t)$.

$$A = \mathcal{T}(A_1, A_2, \ldots, A_t) = \begin{pmatrix} A_1 & A_{12} & A_{13} & \cdots & A_{1t} \\ 0 & A_2 & A_{23} & \cdots & A_{2t} \\ 0 & 0 & A_3 & \cdots & A_{3t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_t \end{pmatrix}.$$

If A_i and B_i have the same size for each i, then the product of $A = \mathcal{T}(A_1, A_2, \ldots, A_t)$ with $B = \mathcal{T}(B_1, B_2, \ldots, B_t)$ has the form $\mathcal{T}(A_1B_1, A_2B_2, \ldots, A_tB_t)$. When all the off-diagonal blocks are zero ($A_{ij} = 0$ for $i \neq j$) then we say A is *block diagonal*, and say A is $\mathcal{D}(A_1, A_2, \ldots, A_t)$ or write $A = \mathcal{D}(A_1, A_2, \ldots, A_t)$. We also say A is the *direct sum* of $A_1, A_2, \ldots, A_t$.

We use $N(A)$ to denote the null space of A and null(A) for the nullity of A, i.e., the dimension of $N(A)$. The range space of A is denoted by $R(A)$ and rank(A) denotes the rank of A, i.e., the dimension of $R(A)$.

We use I_k to denote the $k \times k$ identity matrix and 0_k for the $k \times k$ zero matrix. For $r > s$, the notation $I_{r,s}$ means a matrix with r rows and s columns in which the first s rows are I_s and the remaining $r - s$ rows are rows of zeroes. For

example,

$$I_{5,3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

A matrix with linearly independent columns is said to have *full column rank*; for example $I_{r,s}$ has full column rank. Note that if $r > s$ and A is an $r \times s$ matrix with full column rank, then there exists a nonsingular $r \times r$ matrix B such that $BA = I_{r,s}$.

3. REDUCTION TO THE NILPOTENT CASE. As with the Jordan form, deriving the Weyr form boils down to analyzing the action of the linear transformation on its generalized eigenspaces, and ultimately to analyzing nilpotent transformations.

Let T be a linear operator on V, and let spec$(T) = \{\alpha_1, \alpha_2, \dots, \alpha_t\}$ denote the set of distinct eigenvalues, or spectrum, of T. The generalized eigenspace for each eigenvalue α_i of T is the subspace

$$V_{\alpha_i} = \left\{x \in V | (T - \alpha_i I)^k x = 0 \quad \text{for some nonnegative integer } k\right\}.$$

The space V_{α_i} is invariant under T and contains the eigenspace $U_{\alpha_i} = \{x \in V | (T - \alpha_i I)x = 0\}$. Furthermore, V is the direct sum of the generalized eigenspaces V_{α_i}. Thus, setting $V_i = V_{\alpha_i}$, we have $V = V_1 \oplus V_2 \oplus \cdots \oplus V_t$. Now let n_i be the dimension of V_i and let T_i denote the action of T on the subspace V_i. Choose a basis for each V_i and form a basis B for V by taking the union of these bases. Then the matrix of T with respect to B is $\mathscr{D}(n_1, \dots, n_t)$, where the ith diagonal block represents T_i. Thus, we can describe a canonical form for T by describing a form for the blocks, or for each T_i. Now let $N_i = T - \alpha_i I_{n_i}$. Then N_i is a nilpotent linear operator on V_i and we have reduced the problem to analyzing the action of a nilpotent linear operator or matrix.

4. THE WEYR CHARACTERISTIC FOR THE NILPOTENT CASE. Suppose A is an $n \times n$ nilpotent matrix. The smallest positive integer k such that $A^k = 0$ is called the *index* of A. Then

$$N(A) \subsetneqq N(A^2) \subsetneqq N(A^3) \subsetneqq \cdots \subsetneqq N(A^k) = V$$

and so $0 < \text{null}(A) < \text{null}(A^2) < \cdots < \text{null}(A^k) = n$. For $i = 1, \dots, k$, set $\omega_i = \text{null}(A^i) - \text{null}(A^{i-1})$. The sequence of positive numbers $\omega_1, \omega_2, \dots, \omega_k$ is called the *Weyr characteristic* of A; in Lemma 2 we show that the sequence $\omega_1, \omega_2, \dots, \omega_k$ is non-increasing. We write $\omega(A) = (\omega_1, \omega_2, \dots, \omega_k)$. Note that $\omega_1 = \text{null}(A)$.

We begin by showing how to compute $\omega(A)$ via a recursive process that avoids computing the powers of A; lemmas 1 and 2 are based on work of Kublanovskaya **[13]**. If $k = 1$, then A is the zero matrix, so we may safely assume that $k \geq 2$. Since $\omega_1 = \text{null}(A)$, the matrix A is similar to a matrix with zeros in the first ω_1 columns and thus we can assume A is in the block form

$$\begin{pmatrix} 0 & A_{12} \\ 0 & A_2 \end{pmatrix}$$

where A_{12} is $\omega_1 \times (n - \omega_1)$ and A_2 is square of size $n - \omega_1$. If we are working over the complex numbers, A can be transformed to this block form with a unitary

similarity, because we can choose an orthonormal basis for $N(A)$ and can then extend it to an orthonormal basis for the whole space. Since $\operatorname{rank}(A) = n - \omega_1$, the matrix

$$\begin{pmatrix} A_{12} \\ A_2 \end{pmatrix}$$

has linear independent columns.

Lemma 1. *Suppose A is an $n \times n$ matrix in the form $\mathscr{T}(0_{\omega_1}, A_2)$, where $\omega_1 = \operatorname{null}(A)$. Partition X in F^n as*

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

where $X_1 \in F^{\omega_1}$ and $X_2 \in F^{n-\omega_1}$. Then for any given positive integer r, we have $A^r X = 0$ if and only if $A_2^{r-1} X_2 = 0$.

Proof: Since

$$A^r = \begin{pmatrix} 0 & A_{12} A_2^{r-1} \\ 0 & A_2^r \end{pmatrix},$$

we have

$$A^r X = \begin{pmatrix} A_{12} \\ A_2 \end{pmatrix} \left(A_2^{r-1} X_2 \right).$$

Since the rank of A is $n - \omega_1$, the matrix

$$\begin{pmatrix} A_{12} \\ A_2 \end{pmatrix}$$

has linearly independent columns, and so

$$\begin{pmatrix} A_{12} \\ A_2 \end{pmatrix} Y = 0$$

if and only if $Y = 0$. Putting $Y = A_2^{r-1} X_2$ we see that $A^r X = 0$ if and only if $A_2^{r-1} X_2 = 0$. ∎

Lemma 2. *Let $A = \mathscr{T}(0_{\omega_1}, A_2)$ be an $n \times n$, nonzero, nilpotent matrix with Weyr characteristic $\omega(A) = (\omega_1, \omega_2, \dots, \omega_k)$. Then $\omega(A_2) = (\omega_2, \dots, \omega_k)$. Furthermore, $\omega_1 \geq \omega_2 \geq \cdots \geq \omega_k$.*

Proof: Lemma 1 ensures that $\operatorname{null}(A^i) = \omega_1 + \operatorname{null}(A_2^{i-1})$, so for each $i \geq 2$ we have $\operatorname{null}(A_2^{i-1}) - \operatorname{null}(A_2^{i-2}) = \operatorname{null}(A^i) - \operatorname{null}(A^{i-1}) = \omega_i$. Thus, $\omega(A_2) = (\omega_2, \dots, \omega_k)$.

To prove that $\omega_{i+1} \leq \omega_i$ we use induction on k, starting with $k = 2$. Now, $\operatorname{rank}(A) \leq \operatorname{rank}(A_{12}) + \operatorname{rank}(A_2)$. Substituting $\operatorname{rank}(A) = n - \omega_1$ and $\operatorname{rank}(A_2) = (n - \omega_1) - \operatorname{null}(A_2)$ gives $\operatorname{null}(A_2) \leq \operatorname{rank}(A_{12})$. But $\omega_2 = \operatorname{null}(A_2)$ and $\operatorname{rank}(A_{12}) \leq \omega_1$, so $\omega_2 \leq \omega_1$. By the induction hypothesis, the result holds for the matrix A_2 and so we have $\omega_{i+1} \leq \omega_i$ for all $i \geq 2$. ∎

Lemma 2 leads to a recursive process for computing the Weyr characteristic of a nilpotent matrix. First one applies a similarity to put A in the form $\mathscr{T}(0_{\omega_1}, A_2)$, where $\omega_1 = \text{null}(A)$. This is equivalent to finding the null space of A and choosing a basis, B, for V in which the first ω_1 vectors of B are a basis for $N(A)$. When $F = C$, this can be done with a unitary similarity by choosing B to be an orthonormal basis. Lemma 2 tells us that we have now reduced the problem to finding the Weyr characteristic of the smaller matrix A_2. Repeated application of Lemma 2 leads to a block triangular form in which the diagonal blocks are zero blocks of sizes $\omega_1, \omega_2, \dots, \omega_k$. In Section 5 we examine this form more carefully and show that the superdiagonal blocks have full column rank; this leads to the Weyr canonical form.

We now look at the relationship between the Weyr and Segre characteristics of A. Let S_r denote the $r \times r$ matrix with a 1 in each superdiagonal position and zeros elsewhere; S_r is a nilpotent matrix of index r. Observe that as we form powers of S_r, the superdiagonal line of ones moves upwards, and for $1 \leq m \leq r$, the power S_r^m has rank $r - m$ and nullity m. The Jordan canonical form of A is $J = \mathscr{D}(S_{\sigma_1}, S_{\sigma_2}, \dots, S_{\sigma_t})$ where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_t$. The list $(\sigma_1, \sigma_2, \dots, \sigma_t)$ is the Segre characteristic of A. Since each block S_{σ_i} has nullity one, $\text{null}(A) = t$. Hence, if $\omega(A) = (\omega_1, \omega_2, \dots, \omega_k)$, then $\omega_1 = t$ is the number of blocks in the Jordan form of A. The nullity of J^2 exceeds $\text{null}(J)$ by exactly the number of blocks of size at least two, so $\text{null}(J^2) = t +$ (the number of blocks of size 2 or more). But $\text{null}(A^2) = \omega_1 + \omega_2$, so ω_2 is the number of blocks in the Jordan form that have size at least 2. Now if we look at J^3, we see that $\text{null}(J^3)$ exceeds $\text{null}(J^2)$ by exactly the number of blocks in J with size greater than or equal to 3, so ω_3 is the number of blocks in the Jordan form that have size at least 3. In general, computing $\text{null}(J^m)$ shows that ω_m is the number of blocks in the Jordan form that have size at least m. If we regard the Weyr and Segre characteristics as partitions of n, then the Weyr characteristic is the *conjugate partition* of the Segre characteristic, and we can easily derive one from the other. Using a *Ferrers diagram* to represent the partition $\omega(A) = (\omega_1, \omega_2, \dots, \omega_t)$, the number of dots in row i is ω_i, while σ_i is the number of dots in column i. For example, if $\omega(A) = (4, 3, 3, 2, 2, 2, 1)$, then the Segre characteristic for A is $(7, 6, 3, 1)$ and the corresponding Ferrers diagram is shown in Figure 3.

5. THE WEYR CANONICAL FORM FOR THE NILPOTENT CASE.

We now obtain Weyr's canonical form for the nilpotent case. Since two nilpotent matrices have the same Weyr characteristic if and only if they have the same Segre characteristic, we see that two nilpotent matrices are similar if and only if they have the same Weyr characteristic. Now let $W = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$ be the block

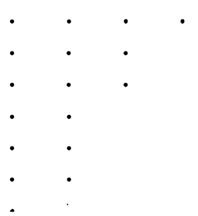


Figure 3. Ferrers diagram for $(4, 3, 3, 2, 2, 2, 1)$.

triangular matrix in which each superdiagonal block is $W_{i,i+1} = I_{\omega_i, \omega_{i+1}}$ and all other blocks are zero. Thus,

$$W = \begin{pmatrix} 0_{\omega_1} & I_{\omega_1,\omega_2} & 0 & \cdots & 0 \\ & 0_{\omega_2} & I_{\omega_2,\omega_3} & \cdots & 0 \\ & & \ddots & & \\ & & & 0_{\omega_{k-1}} & I_{\omega_{k-1},\omega_k} \\ 0 & & & & 0_{\omega_k} \end{pmatrix}.$$

Direct calculation of the powers of W shows that W has Weyr characteristic $(\omega_1, \omega_2, \ldots, \omega_k)$. Hence, W is a canonical form for all nilpotent matrices with Weyr characteristic $(\omega_1, \omega_2, \ldots, \omega_k)$.

This approach is quick and easy, but it depends on the Jordan canonical form. Weyr, of course, developed his theory independently. The remainder of this section presents a derivation of Weyr's form that does not depend on the Jordan canonical form. We use Lemma 2 to obtain a block triangular form $\mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \ldots, 0_{\omega_k})$, show that the superdiagonal blocks have full column rank, and then show how to further reduce this form to obtain the Weyr canonical form. The proofs of the main results are by induction; to get started we need the following lemma.

Lemma 3. *Let T be a nilpotent linear operator on V with $\omega(T) = (\omega_1, \omega_2, \ldots, \omega_k)$. Then T can be represented by a matrix $A = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \hat{A})$, where $\operatorname{rank}(A_{12}) = \omega_2$ and so A_{12} has full column rank.*

Proof: Since $\omega_1 = \operatorname{null}(T)$, we can represent T by a matrix $B = \mathscr{T}(0_{\omega_1}, B_2)$. Lemma 2 ensures that $\omega_2 = \operatorname{null}(B_2)$ so there is a square matrix Q of size $n - \omega_1$ such that $Q^{-1}B_2Q = \mathscr{T}(0_{\omega_2}, \tilde{A})$. Now let $P = \mathscr{D}(I_{\omega_1}, Q)$; then $P^{-1}BP = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \tilde{A})$, so $A = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \tilde{A})$ is a matrix representation for T:

$$\mathscr{T}\left(0_{\omega_1}, 0_{\omega_2}, \tilde{A}\right) = \begin{pmatrix} 0_{\omega_1} & A_{12} & A_{13} \\ 0 & 0_{\omega_2} & A_{23} \\ 0 & 0 & \tilde{A} \end{pmatrix}.$$

Since A has rank $n - \omega_1$, the last $n - \omega_1$ columns of A must be linearly independent, and hence the block A_{12} (which is $\omega_1 \times \omega_2$) must have full column rank. ∎

When $k = 2$, Lemma 3 tells us that T can be represented by a block triangular matrix $\mathscr{T}(0_{\omega_1}, 0_{\omega_2})$, where the $\omega_1 \times \omega_2$ block A_{12} has full column rank, i.e., $\operatorname{rank}(A_{12}) = \omega_2$.

Remark 1. If $F = C$, then in the proof of Lemma 3, we can use an orthonormal basis for C^n in which the first ω_1 vectors are a basis for $N(T)$ and can use a unitary matrix for Q. Hence, we can obtain a representation for T in the form given in Lemma 3 by using an appropriate orthonormal basis.

Theorem 1. *Let T be a nilpotent linear operator on V. Then $\omega(T) = (\omega_1, \omega_2, \ldots, \omega_k)$ if and only if T can be represented by a block triangular matrix $A = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \ldots, 0_{\omega_k})$*

in which each superdiagonal block $A_{i,i+1}$ has full column rank, i.e., rank$(A_{i,i+1}) = \omega_{i+1}$.

Proof: We use induction on k. Assume $\omega(T) = (\omega_1, \omega_2, \dots, \omega_k)$. If $k = 1$, then T is the zero matrix. If $k = 2$, then Lemma 3 gives the result. For the general case, we apply Lemma 3 to see that T has a matrix representation $B = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \tilde{B})$, where B_{12} has full column rank. Let B_2 denote the square submatrix in the last $n - \omega_1$ rows and columns; then B_2 is $\mathscr{T}(0_{\omega_2}, \tilde{B})$. Lemma 2 tells us that $\omega(B_2) = (\omega_2, \dots, \omega_k)$, so by the induction hypothesis, there is a nonsingular matrix Q, of size $n - \omega_1$, such that $Q^{-1}B_2Q = \mathscr{T}(0_{\omega_2}, 0_{\omega_3}, \dots, 0_{\omega_k})$ with each superdiagonal block having full column rank. Apply the similarity $P = \mathscr{D}(I_{\omega_1}, Q)$ to B to get a matrix, A, in the desired form.

To prove the converse, it suffices to show that a matrix $A = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$ with superdiagonal blocks of full column rank has Weyr characteristic $(\omega_1, \omega_2, \dots, \omega_k)$. We again use induction on k. Observe that the last $n - \omega_1$ columns of such a matrix are linearly independent, so null$(A) = \omega_1$. If $k = 1$, then A is the zero matrix and we are done. Otherwise, A has the form $\mathscr{T}(0_{\omega_1}, A_2)$ given in Lemma 1, and Lemma 2 tells us that the Weyr characteristic of A is $(\omega_1, \omega_2', \dots, \omega_k')$, where $(\omega_2', \dots, \omega_k') = \omega(A_2)$. But the induction hypothesis then tells us that $\omega_i = \omega_i'$ for $i \geq 2$ and we are done. ■

Using Remark 1 and a unitary matrix for the matrix Q in the proof of Theorem 1, we obtain the following unitary version of Theorem 1.

Theorem 1′. *Let A be an $n \times n$ nilpotent complex matrix. Then $\omega(A) = (\omega_1, \omega_2, \dots, \omega_k)$ if and only if there is a unitary matrix U such that U^*AU is a block triangular matrix of the form $\mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$ in which each superdiagonal block $A_{i,i+1}$ has full column rank, i.e., rank$(A_{i,i+1}) = \omega_{i+1}$.*

It is also possible to apply further unitary similarities to reduce the superdiagonal blocks to special forms; see **[2]**, **[21]**, and **[22]**.

For purposes of computing the Weyr characteristic, one would stop with the staircase form of Theorem 1′, which can be reached via a unitary similarity. However, this block triangular form is not unique; for a canonical form we must go further and use non-unitary similarities.

Theorem 2. *Let T be a nilpotent linear operator on V. Then $\omega(T) = (\omega_1, \omega_2, \dots, \omega_k)$ if and only if T can be represented by the block triangular matrix $W = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$, in which the only nonzero blocks are the superdiagonal blocks $W_{i,i+1} = I_{\omega_i, \omega_{i+1}}$, $i = 1, \dots, k - 1$.*

Proof: Using Theorem 1, it suffices to show that a matrix $B = \mathscr{T}(0_{\omega_1}, 0_{\omega_2}, \dots, 0_{\omega_k})$ in which each superdiagonal block has full column rank is similar to W. We use induction on k. When $k = 1$, we have $B = 0$ and there is nothing to do. Assume $k > 1$. The matrix occupying the last $n - \omega_1$ rows and columns of B has Weyr characteristic $(\omega_2, \omega_3, \dots, \omega_k)$, so the induction hypothesis ensures that it is similar to a matrix in the desired form. Thus, there is a nonsingular matrix Q, of size $n - \omega_1$, such that $C = \mathscr{D}(I_{\omega_1}, Q^{-1})B\mathscr{D}(I_{\omega_1}, Q)$ has the desired form except

possibly in the first row of blocks, $(0_{\omega_1}, C_{12}, C_{13}, \ldots, C_{1k})$. Thus,

$$C = \begin{pmatrix} 0_{\omega_1} & C_{12} & C_{13} & C_{14} & \cdots & C_{1k} \\ & 0_{\omega_2} & I_{\omega_2,\,\omega_3} & 0 & \cdots & 0 \\ & & 0_{\omega_3} & I_{\omega_3,\,\omega_4} & \cdots & 0 \\ & & & \ddots & & \\ & & & & & I_{\omega_{k-1},\,\omega_k} \\ 0 & & & & & 0_{\omega_k} \end{pmatrix}.$$

Now, $\text{null}(B) = \text{null}(C) = \omega_1$ so C_{12} has full column rank. We now reduce C to the desired form in two steps. First, we clear out the blocks $C_{13}, \ldots, C_{1k}$, and then we reduce C_{12} to the form $I_{\omega_1,\,\omega_2}$.

The block C_{1r} is $\omega_1 \times \omega_r$; let $\tilde{C}_{1r}$ denote the $\omega_1 \times \omega_{r-1}$ matrix obtained by adjoining $\omega_{r-1} - \omega_r$ columns of zeros to C_{1r}. Thus, we have

$$\tilde{C}_{1r} = \begin{pmatrix} C_{1r} & 0_{\omega_1 \times (\omega_{r-1} - \omega_\omega)} \end{pmatrix},$$

and $\tilde{C}_{1r} I_{\omega_{r-1},\,\omega_r} = C_{1r}$. Now let P be the matrix of the form $\mathscr{T}(I_{\omega_1}, I_{n-\omega_1})$ in which the first ω_1 rows are the blocks $(I_{\omega_1}, \tilde{C}_{13}, \tilde{C}_{14}, \ldots, \tilde{C}_{1k}, 0_{\omega_1 \times \omega_k})$, that is,

$$P = \left(\begin{array}{c|c} I_{\omega_1} & \tilde{C}_{13}\tilde{C}_{14} \cdots \tilde{C}_{1k}\, 0_{\omega_1 \times \omega_k} \\ \hline & I_{n-\omega_1} \end{array}\right).$$

Then P^{-1} has the same form, but its first ω_1 rows are the blocks $(I_{\omega_1}, -\tilde{C}_{13}, -\tilde{C}_{14}, \ldots, -\tilde{C}_{1k}, 0_{\omega_1 \times \omega_k})$. A computation using block multiplication shows that $P^{-1}CP$ has C_{12} in its $1,2$ block, but otherwise has the desired form.

Since C_{12} has full column rank, there is a nonsingular $\omega_1 \times \omega_1$ matrix W such that $WC_{12} = I_{\omega_1,\,\omega_2}$. Let $S = \mathscr{D}(W^{-1}, I_{\omega_2}, I_{\omega_3}, \ldots, I_{\omega_k})$; then $S^{-1}P^{-1}CPS$ has the desired form. ∎

6. THE GENERAL CASE. We can now use our form for the nilpotent case to deal with a general linear operator T. As described in Section 2, we can decompose T into a direct sum $T_1 \oplus T_2 \oplus \cdots T_t$, where each T_i is the action of T on the generalized eigenspace V_i. Then $T_i - \alpha_i I$ is a nilpotent transformation on V_i. We say that $\omega(T_i - \alpha_i I)$ is the *Weyr characteristic of T, relative to the eigenvalue α_i*. Let W_i be the Weyr canonical form of N_i; then T can be represented by the block diagonal matrix $\mathscr{D}(\alpha_1 I + W_1, \alpha_2 I + W_2, \ldots, \alpha_t I + W_t)$. This is the canonical form described by Weyr [**28**]; we call it the *Weyr canonical form* of T. For each eigenvalue, α_i, the Weyr characteristic, $\omega(T_i - \alpha_i I)$ is related to the Segre characteristic for α_i as described in Section 4, and so the Jordan canonical form of a matrix can be read off from the Weyr canonical form, and vice versa.

7. OBTAINING THE WEYR CHARACTERISTIC BY UNITARY SIMILARITY. Two $n \times n$ complex matrices, A and B, are *unitarily similar* if there is a unitary matrix U such that $B = U^*AU$. In general, a matrix is not unitarily similar to its Jordan or Weyr canonical form. However, in numerical computations, it is desirable to obtain the information needed to specify the canonical form by using only unitary similarities. We briefly outline, in theory, why the Weyr characteristic can be found using only unitary similarities.

The process begins with Schur's result that a square complex matrix can be triangularized with a unitary similarity [**11**, pp. 79–81].

Theorem (Schur [**20**]). *If A is an $n \times n$ complex matrix, then there is a unitary matrix U such that U^*AU is triangular.*

Proof: Start with an eigenvalue, α_1, of A and an associated eigenvector x, where x has length one. Then construct an orthonormal basis for C^n in which x is the first basis element. Let U_1 be the unitary matrix that has the basis vectors in its columns. Then $U_1^*AU_1$ has the form $\mathscr{T}(\alpha_1, A_1)$ where A_1 is $(n-1) \times (n-1)$. Using induction, let U_2 be a unitary matrix of size $n-1$ that puts A_1 in triangular form and let $U_2 = \mathscr{D}(1, \tilde{U}_2)$. Then if $U = U_1U_2$, the matrix U^*AU is triangular. ■

Note that we can obtain a triangular form for A with the eigenvalues in any given order along the diagonal. Thus, if $\text{spec}(A) = \{\alpha_1, \alpha_2, \ldots, \alpha_t\}$, where α_i has multiplicity n_i, we can unitarily put A into the form $\mathscr{T}(A_1, A_2, \ldots, A_t)$ where A_i is an $n_i \times n_i$ triangular matrix with α_i along its diagonal.

The next step is to show that $\mathscr{T}(A_1, A_2, \ldots, A_t)$ is similar to $\mathscr{D}(A_1, A_2, \ldots, A_t)$, for this will tell us that the Weyr characteristic of A, relative to the eigenvalue α_i is simply the Weyr characteristic of the nilpotent matrix $A_i - \alpha_i I$. To show that $\mathscr{T}(A_1, A_2, \ldots, A_t)$ and $\mathscr{D}(a_1, A_2, \ldots, A_t)$ are similar, we use a well-known theorem of Sylvester, which may be found in many sources, e.g., [**3**], [**6**, Vol 1, p. 225], [**11**, Section 2.4, Problems 9 and 13], and [**12**, Theorem 4.4.6].

Theorem (Sylvester) [**23**]. *Let A be $m \times m$ and B be $n \times n$. Then the matrix equation $AX - XB = C$ has a unique solution for every $m \times n$ matrix C if and only if $\text{spec}(A) \cap \text{spec}(B) = \varnothing$.*

Lemma 4. *If $A = \mathscr{T}(A_1, A_2)$ and $\text{spec}(A_1) \cap \text{spec}(A_2) = \varnothing$ then A is similar to $\mathscr{D}(A_1, A_2)$.*

Proof: Let A_i be size $n_i \times n_i$ for $i = 1, 2$. Let X be the unique $n_1 \times n_2$ matrix that satisfies $A_1X - XA_2 = -A_{12}$. Let S be of the form $\mathscr{T}(I_{n_1}, I_{n_2})$ with X in the 1, 2 block. Then S^{-1} is $\mathscr{T}(I_{n_1}, I_{n_2})$ with $-X$ in the 1, 2 block. A computation then shows that $S^{-1}AS$ is $\mathscr{D}(A_1, A_2)$.

Using Lemma 4 with an induction argument proves the following result.

Theorem 3. *If $A = \mathscr{T}(A_1, A_2, \ldots, A_t)$, where each $\text{spec}(A_i) = \{\alpha_i\}$ and $\alpha_i \neq \alpha_j$ when $i \neq j$, the A is similar to $\mathscr{D}(A_1, A_2, \ldots, A_t)$.*

Thus, once we have A in the triangular form $\mathscr{T}(A_1, A_2, \ldots, A_t)$, we can find the Weyr characteristic of each eigenvalue of A by finding the Weyr characteristic of each nilpotent block $A_i - \alpha_i I$. As pointed out in Section 4, this can be done with a recursive procedure and can be done with unitary transformations. We refer the reader to references [**8**], [**13**], and [**18**] for detailed information on numerical algorithms and the stability issues involved.

REFERENCES

1. Th. Beelen and P. Van Dooren, An improved algorithm for the computation of Kronecker's canonical form for a singular pencil, *Linear Algebra Appl.* **105** (1988) 9–65.
2. R. Benedetti and P. Cragnolini, Versal families of matrices with respect to unitary conjugation, *Adv. Math.* **54** (1984) 314–335.
3. R. Bhatia and P. Rosenthal, How and why to solve the operator equation $AX - XB = Y$, *Bull. London Math. Soc.* **29** (1997) 1–21.

4. Richard Brualdi, The Jordan Canonical Form: an Old Proof, *Amer. Math. Monthly* **94** (1987) 257–267.
5. J. W. Demmel, *Applied Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, 1997.
6. F. R. Gantmacher, *The Theory of Matrices*, Vols. 1, 2, Chelsea, New York, 1959.
7. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd edition, The Johns Hopkins University Press, Baltimore and London, 1989.
8. G. H. Golub and J. H. Wilkinson, Ill-conditioned eigensystems and the computation of the Jordan canonical form, *SIAM Review* **18** (1976) 578–619.
9. D. Hershkowitz and H. Schneider, On the existence of matrices with prescribed height and level characteristics, *Israel J. Math.* **75** (1991) 105–117.
10. D. Hershkowitz and H. Schneider, Height bases, level bases, and the equality of the height and the level characteristics of an M-matrix, *Linear and Multilinear Algebra*, **25** (1989) 149–171.
11. R. Horn and C, Johnson, *Matrix Analysis*, Cambridge U. P., Cambridge, 1985.
12. R. Horn and C. Johnson, *Topics in Matrix Analysis*, Cambridge U. P., Cambridge, 1990.
13. V. N. Kublanovskaya, On a method of solving the complete eigenvalue problem for a degenerate matrix, *U.S.S.R. Comput. Math. and Math. Physics* **6** (1966) 1–14.
14. D. E. Littlewood, On unitary equivalence, *J. London Math. Soc.* **28** (1953) 314–322.
15. C. C. MacDuffee, *The Theory of Matrices*, Springer Verlag, Berlin, 1933.
16. A. I. Mal'cev, *Foundations of Linear Algebra*, W. H. Freeman and Company, San Francisco and London, 1963.
17. D. J. Richman and H. Schneider, On the singular graph and the Weyr characteristic of an M-matrix, *Aequationes Math.* **17** (1978) 208–234.
18. A. Ruhe, An algorithm for numerical determination of the structure of a general matrix, *BIT* 10 (1970) 196–216.
19. H. Schneider, The influence of the marked reduced graph of a nonnegative matrix on the Jordan form and related properties: A survey, *Linear Algebra Appl.* **84** (1986) 161–189.
20. I. Schur, Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen, *Math. Ann.* **66** (1909) 488–510.
21. V. V. Sergeichuk, Classification of linear operators in a finite-dimensional unitary space, *Functional Anal. Appl.* **18** (1984) 224–230.
22. H. Shapiro, A survey of canonical forms and invariants for unitary similarity, *Linear Algebra Appl.* **147** (1991) 101–167.
23. J. J. Sylvester, Sur l'equation en matrices $px = xq$, *C. R. Acad. Sci. Paris* **99** (1884) 67–71 and 115–116.
24. H. W. Turnbull and A. C. Aitken, *An Introduction to the theory of Canonical Matrices*, Blackie & Son Limited, London and Glasgow, 1932.
25. P. Van Dooren, *The Generalized Eigenstructure Problem*; *Applications in Linear System Theory*, Ph.D. Thesis, Katholieke Universiteit Leuven, May, 1979.
26. P. Van Dooren, The computation of Kronecker's canonical form of a singular pencil, *Linear Algebra Appl.* **27** (1979) 103–140.
27. P. Van Dooren, The generalized eigenstructure problem in linear system theory, *IEEE Trans. Automatic Control* **26** (1981) 111–129.
28. E. Weyr, Zur Theorie der bilinearen Formen, *Monatsh. Math. und Physik* **1** (1890) 163–236.
29. E. Weyr, Répartition des matrices en espèces et formation de toutes les espèces, *C. R. Acad. Sci. Paris* **100** (1885) 966–969.

HELENE SHAPIRO received her B.A. degree from Kenyon College in 1975 and her Ph.D. from the California Institute of Technology in 1979, where she was a student of Olga Taussky Todd. She spent one year at the University of Wisconsin and then joined the Department of Mathematics and Statistics at Swarthmore College, where she has been teaching since 1980.
Department of Mathematics and Statistics, Swarthmore College, Swarthmore, PA 19081
hshapir1@swarthmore.edu

Rental Harmony: Sperner's Lemma in Fair Division

Francis Edward Su

My friend's dilemma was a practical question that mathematics could answer, both elegantly and constructively. He and his housemates were moving to a house with rooms of various sizes and features, and were having trouble deciding who should get which room and for what part of the total rent. He asked, "Do you think there's always a way to partition the rent so that each person prefers a different room?"

As we shall see, with mild assumptions, the answer is yes. This rent-partitioning problem is really a kind of *fair-division* question. It can be viewed as a generalization of the age-old *cake-cutting* problem, in which one seeks to divide a cake fairly among several people, and the *chore-division* problem, posed by Martin Gardner in [**6**, p. 124], in which one seeks to fairly divide an undesirable entity, such as a list of chores. Lately, there has been much interest in fair division (see the recent books [**3**] and [**11**]), and each of the related problems has been treated before (see [**1**], [**4**], [**10**]).

We wish to explain a powerful approach to fair-division questions that unifies these problems and provides new methods for achieving approximate *envy-free* divisions, in which each person feels she received the "best" share. This approach was carried out by Forest Simmons [**12**] for cake-cutting and depends on a simple combinatorial result known as Sperner's lemma. We show that the Sperner's lemma approach can be adapted to treat chore division and rent-partitioning as well, and it generalizes easily to any number of players.

From a pedagogical perspective, this approach provides a nice, elementary demonstration of how ideas from many pure disciplines—combinatorics, topology, and analysis—can combine to address a real-world problem. Better yet, the proofs can be converted into constructive fair-division procedures.

1. SPERNER'S LEMMA FOR TRIANGLES. Our fair division approach is based on a simple combinatorial lemma, due to Sperner [**13**] in 1928. However, do not be fooled—this little lemma is as powerful as it is simple. It can, for instance, be used to give a short, elementary proof of the Brouwer fixed point theorem [**7**].

As motivation, we examine a special case of Sperner's lemma. Consider a triangle T triangulated into many smaller triangles, called *elementary* triangles, whose vertices are labelled by 1's, 2's, and 3's, as in Figure 1.

The labelling we have chosen obeys two conditions: (1) all of the main vertices of T have different labels, and (2) the label of a vertex along any edge of T matches the label of one of the main vertices spanning that edge; labels in the interior of T are arbitrary. Any labelled triangulation of T satisfying these conditions is called a *Sperner labelling*. The claim:

Sperner's Lemma for Triangles. *Any Sperner-labelled triangulation of T must contain an odd number of elementary triangles possessing all labels. In particular, there is at least one.*

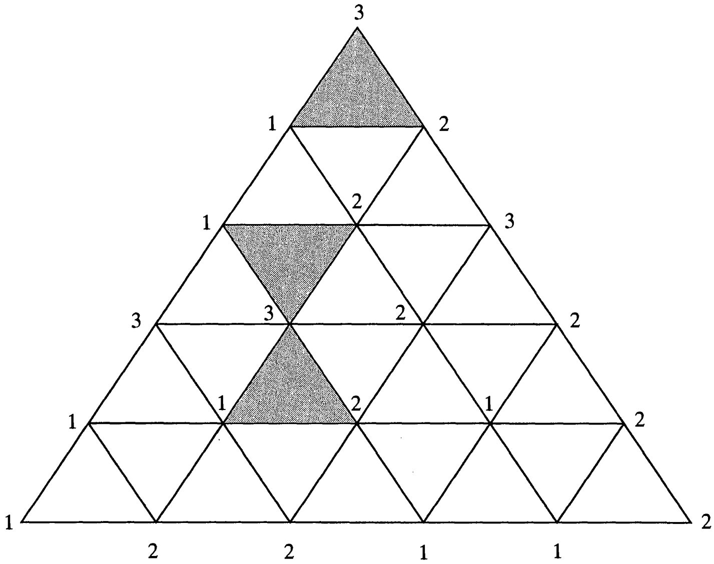


Figure 1. A Sperner labelling, with (1,2,3)-triangles marked.

In Figure 1, we have marked all elementary 123-triangles; their parity is indeed odd. An analogous statement holds in any dimension, which we develop presently.

2. THE n-DIMENSIONAL SPERNER'S LEMMA. We need the concept of an *n-simplex*: a 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a triangle, a 3-simplex is a tetrahedron, etc. In general, an *n-simplex* may be regarded as an n-dimensional "tetrahedron"—the convex hull of $n + 1$ affinely independent points in $\mathbf{R}^m$, for $m \geq n$. These points form the vertices of the simplex. A *k-face* of an n-simplex is the k-simplex formed by the span of any subset of $k + 1$ vertices.

A *triangulation* of an n-simplex S is a collection of (distinct) smaller n-simplices whose union is S, with the property that any two of them intersect in a face common to both, or not at all. The smaller n-simplices are called *elementary simplices*, and their vertices are called *vertices of the triangulation.*

Given an n-simplex S, any face spanned by n of the $n + 1$ vertices of S is called a *facet*. As examples, the facets of a line segment are its endpoints, the facets of a triangle are its sides, and the facets of a tetrahedron are its triangular faces.

Now number the facets of S by $1, 2, \ldots, n + 1$. Given a triangulation of S, consider a labelling that obeys the following rule: each vertex is labelled by one of the facet numbers in such a way that on the boundary of S, none of the vertices on facet j is labelled j. The interior vertices can be labeled by any of the facet numbers. Such a labelling is called a *Sperner labelling of an n-simplex*; it generalizes the definition we encountered earlier for $n = 2$. For other low dimensions, Figures 2 and 3 show examples of a Sperner-labelled 1-simplex and 3-simplex.

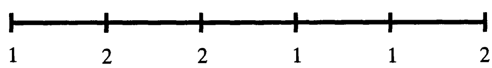


Figure 2. A triangulated line, with Sperner labelling.

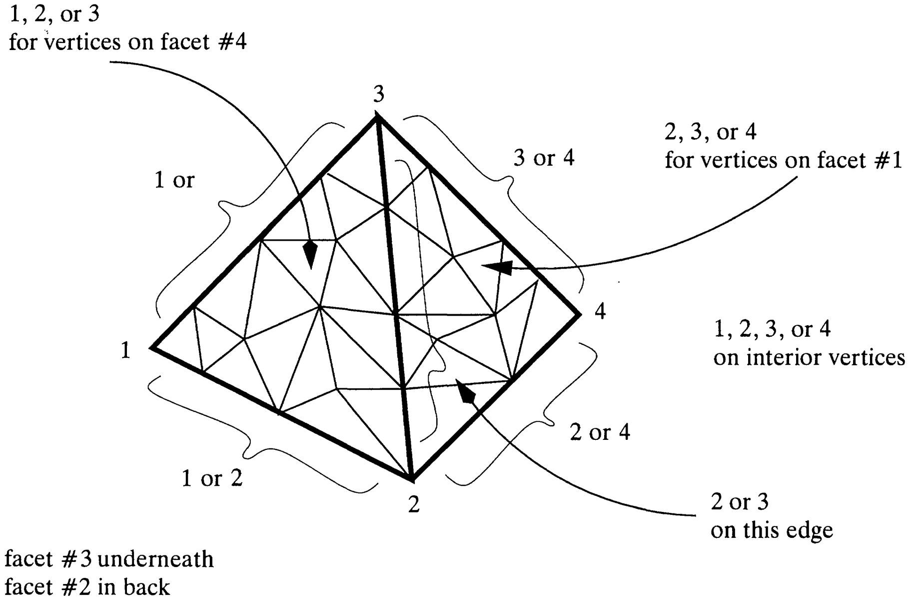


Figure 3. A triangulated tetrahedron, with Sperner labelling.

A Sperner labelling may be described equivalently as one in which main vertices of S are assigned distinct labels, and any other vertex in the interior of some k-face must be assigned one of the labels of the main vertices that span that face. In either description it is apparent that the Sperner labelling on S induces Sperner labellings on each facet as $(n - 1)$-simplices.

We call an elementary simplex in the triangulation *fully labelled* if all its vertices have distinct labels. Then we have:

Sperner's Lemma. *Any Sperner-labelled triangulation of a n-simplex must contain an odd number of fully labelled elementary n-simplices. In particular, there is at least one.*

There are many ways to prove this lemma. The simplest proofs involve parity arguments and are non-constructive. A constructive method for finding a fully labelled simplex is based on the following induction argument; it is useful later in our discussion of fair-division procedures in Sections 5 and 7.

Proof: We proceed by induction on the dimension n.

When $n = 1$, a triangulated 1-simplex is a segmented line, as in Figure 2. The endpoints of the line are labelled distinctly, by 1 and 2. Hence in moving from endpoint 1 to endpoint 2 the labelling must switch an odd number of times, i.e., an odd number of $(1, 2)$-edges may be located in this way.

Now assume that the theorem holds for dimensions up through $(n - 1)$. We show the theorem is true for a triangulated, Sperner-labelled n-simplex S using the labels 1 through $(n + 1)$. For concreteness refer to the case $n = 2$ as a running example while following the argument. In this case, S is a triangulated triangle, as in Figure 4.

Think of the n-simplex S as a "house" triangulated into many "rooms," which are the elementary simplices. A facet of a room is called a "door" if that facet

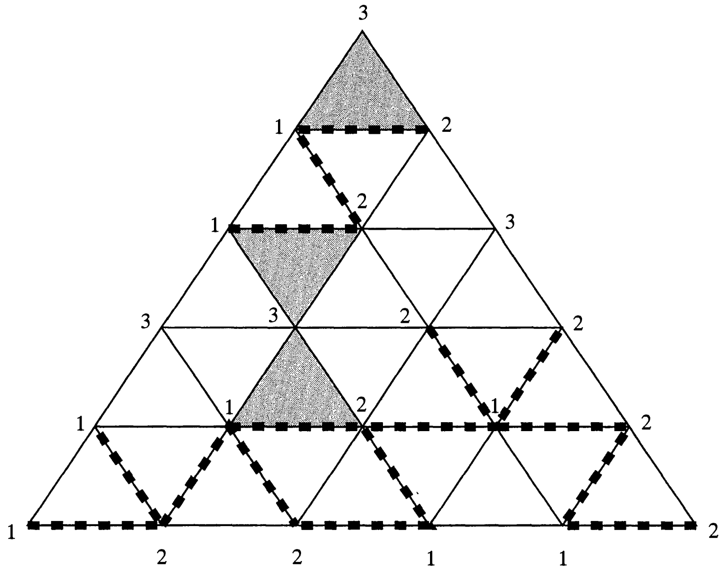


Figure 4. House, rooms, and doors indicated by dotted lines.

carries the first n of the $n+1$ labels. In our running example, doors are $(1,2)$-edges that may be in the interior or on the boundary; see Figure 4. For the case $n=3$, doors are any room facets labelled $(1,2,3)$.

We claim that the number of doors on the boundary of S is odd. Why? The only facet that can contain doors is the $(n+1)$-st because of the Sperner labelling. But that facet of S is Sperner-labelled using the labels $1,\ldots,n$, hence by the inductive hypothesis there must be odd number of fully labelled $(n-1)$-simplices on that facet. These are boundary doors when considered in S.

The boundary doors can be used to locate fully labelled rooms by what we fondly call a "trap-door" argument. The key observation is that every room can have at most 2 doors, and it has exactly 1 door if and only if the room is fully labelled in S. This is true because any room with at least one door has either no repeated labels (it is fully labelled), or it has one repeated label that appears twice. These give rise to 2 distinct doors, one for each repeated label. As examples, verify that elementary triangles in Figure 4 have either two, one, or no $(1,2)$-edges. For $n=3$, verify that a tetrahedron with labels $\{1,2,3,3\}$ has two doors.

So, start at any door on the boundary (located by the inductive step), and "walk" through the door into the adjoining room. Either this room is fully labelled or it has one other door—a "trap-door" that we can walk through. Repeat this procedure, walking through doors whenever possible. Notice that this path cannot double back on itself (because each room has at most two doors), so no room is ever visited twice. Moreover the number of rooms is finite and so the procedure must end, either by walking into a fully labelled room or by walking back through to a boundary door of S; see Figure 5.

Since the number of boundary doors of S is odd, and trap-door paths pair up only an even number of them, the number of boundary doors left over that lead to fully labelled rooms must be odd. Moreover, any fully labelled rooms not reachable by paths from the boundary must come in pairs, matched up by their own trap-door paths, as in Figure 5. Hence the total number of fully labelled rooms in S is odd. ∎

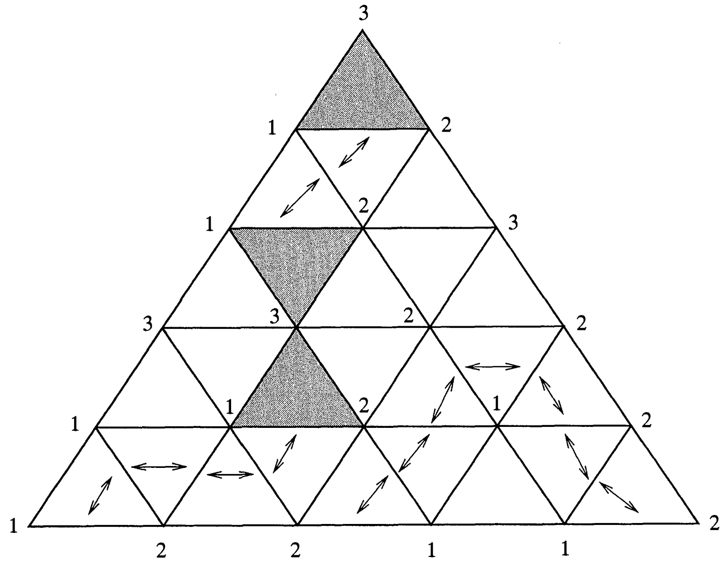


Figure 5. Walking through doors.

This proof yields a constructive method for finding such rooms in the following way. Trap-door paths in successive dimensions can be linked up at their endpoints, because a fully labelled room in an i-dimensional face is just a boundary door in an $(i + 1)$-face. This creates "super-paths" with endpoints in the bottom and top dimensions, i.e., either $(1, 2)$-edges on a 1-face of S, or n-dimensional fully labelled rooms in the interior. The constructive procedure begins by moving along the 1-face of S spanned by labels 1 and 2, following any super-path that is encountered. Because the number of $(1, 2)$-edges is odd, and super-paths can pair up only an even number of them, we see that at least one super-path can be followed to yield a fully labelled room.

The trap-door argument to prove Sperner's lemma constructively dates back to Cohen [**5**] and Kuhn [**8**]. A quick non-constructive proof would note the equality between the number of doors in each room, summed over all rooms, and the number of times each door is counted, summed over all doors. Modulo two, the first sum captures the parity of the number of fully labelled rooms, and the second sum captures the parity of the number of boundary doors, which by the inductive hypothesis is odd.

3. SIMMONS' APPROACH TO CAKE-CUTTING. Now imagine a rectangular cake to be divided among n people, who may have differing notions of what is valuable on a cake. We use $n - 1$ knives to cut along planes parallel to the left edge of the cake, as in Figure 6.

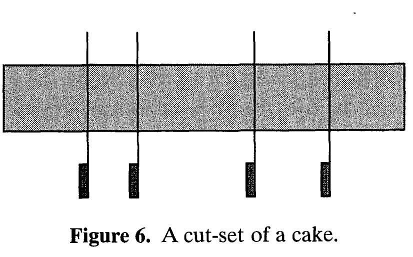


Figure 6. A cut-set of a cake.

The set of cuts is fully defined by the relative sizes of the pieces. Assume that the total size of the cake is 1 and denote the physical size of the i-th piece by x_i; this is an absolute measure, unrelated to player preferences. Thus $x_1 + x_2 + \cdots + x_n = 1$ and each $x_i \geq 0$. The space S of possible partitions naturally forms a standard $(n-1)$-simplex in $\mathbf{R}^n$. Each point in S corresponds to a partition of the cake by a set of cuts, which we shall call a *cut-set*.

Given a cut-set, we say that a player *prefers* a given piece if the player does not think any other piece is better. We assume that this preference depends on the player and the entire cut-set, but not on choices made by the other players. Note that, given a cut-set, a player always prefers at least one piece, and may (in case of ties) prefer more than one piece by our definition.

We make the following two assumptions:

(1) *The players are hungry*. That is, players prefer any piece with mass to an empty piece.
(2) *Preference sets are closed*. This means that any piece that is preferred for a convergent sequence of cut-sets is preferred at the limiting cut-set. Note that this condition rules out the existence of single points of cake with positive desirability.

Theorem. *For hungry players with closed preference sets, there exists an envy-free cake division, i.e., a cut-set for which each person prefers a different piece.*

We first investigate what happens for $n = 3$ people. Suppose the players are named Alice, Betty, and Charlie. They are to divide a cake of total size 1, using 2 knives. Denote the physical size of the pieces by x_1, x_2, x_3. Since $x_1 + x_2 + x_3 = 1$ and all $x_i \geq 0$, the solution space S is a plane intersected with the first octant. This is just a triangle.

Now triangulate S and assign "ownership" to each of the vertices as in Figure 7, where A stands for Alice, B for Betty, and C for Charlie. We have purposely

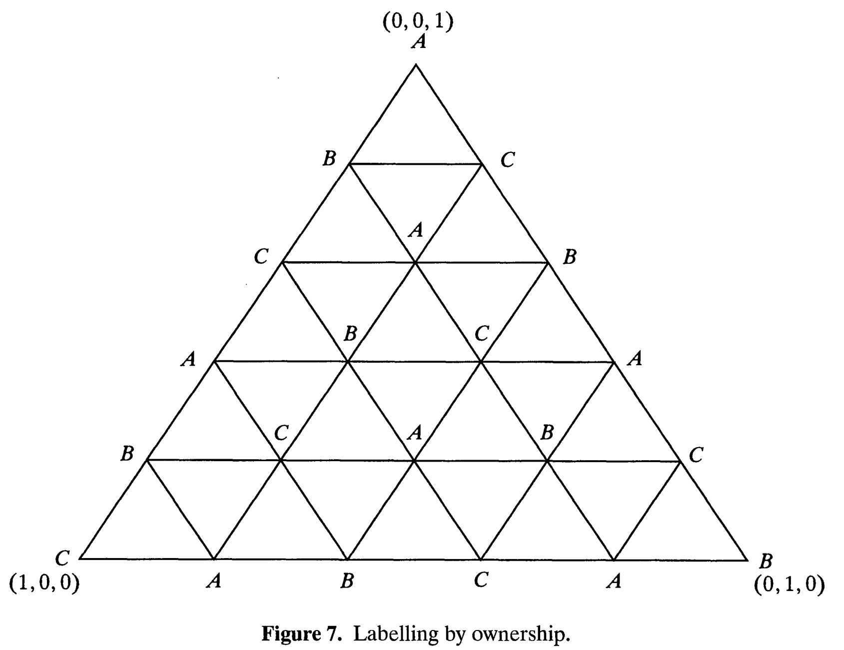


Figure 7. Labelling by ownership.

assigned ownership so that each elementary triangle is an ABC triangle. Observe that a similar triangulation of finer mesh can also be labelled in this way.

We obtain a *new* auxiliary labelling of the triangulation by 1's, 2's, and 3's by doing the following: since each point in the triangle corresponds to a set of cuts of cake, go to each vertex, and ask the owner of that vertex, "Which piece would you choose if the cake were cut with this cut-set?" Label that vertex by the number of the piece that is desired.

We claim that this new labelling is a Sperner labelling! Why?

At the vertex (1,0,0) of S we see that one of the pieces contains the entire cake, and the other pieces are empty. By the hungry assumption, the owner of (1,0,0) always chooses piece 1 no matter who the owner is. Similarly (0,1,0) is labelled 2 and (0,0,1) is labelled 3. Next, observe that the sides of the triangle correspond to cuts in which one piece is devoid of any cake. Because no one would ever choose this empty piece, each side of S is missing one label corresponding to the piece that is empty. Hence the Sperner labelling condition is satisfied.

By Sperner's lemma, there must be a $(1, 2, 3)$-elementary simplex in the triangulation. Since every such simplex arose from an ABC triangle, this means that we have found 3 very similar cut-sets in which different people choose different pieces of cake.

To show the existence of a single cut-set that would satisfy everyone with different pieces, carry out this procedure for a sequence of finer and finer triangulations, each time yielding smaller and smaller $(1, 2, 3)$-triangles. By compactness of the triangle and decreasing size of the triangles, there must be a convergent subsequence of triangles converging to a single point. Such a point corresponds to a cut-set in which the players are satisfied with different pieces. Why?

Since each $(1, 2, 3)$-triangle in the convergent subsequence arises from an ABC triangle, consider the choices that the players made in each. With only finitely many ways for players to choose pieces, there must be an infinite subsequence in which the choices of A, B, and C are all constant. Closed preference sets guarantee that at the limit point of this subsequence of triangles, the players are satisfied with distinct choices.

4. THE n-PLAYER CASE. The preceding proof generalizes easily for n players. The only issue that must be addressed is the choice of triangulation for S when $n > 3$. We need a triangulation in which each elementary simplex can be fully labelled by the names of the players. The triangulation we proposed for $n = 3$ does not generalize easily. However, one that works for arbitrary dimensions is a triangulation by *barycentric subdivision.* Loosely speaking, this procedure takes each elementary simplex in a triangulation and subdivides it by marking the barycenters of the faces in each dimension and connecting them to form a new triangulation. A rigorous description of this procedure may be found in **[15]**. Observe that the mesh of this triangulation can be made arbitrarily small by iterating this procedure; see Figure 8.

Suppose we have iterated barycentric subdivision m times. The desired labelling can be achieved by allowing all vertices that remain from the $(m - 1)$-th iteration to be labelled A. Any new vertices introduced in the m-th barycentric subdivision are barycenters of simplices of the $(m - 1)$-th subdivision. To each class of vertices that are barycenters of faces of the same dimension, assign a distinct owner from the persons remaining. There are $n - 1$ such classes. One may verify that this fully

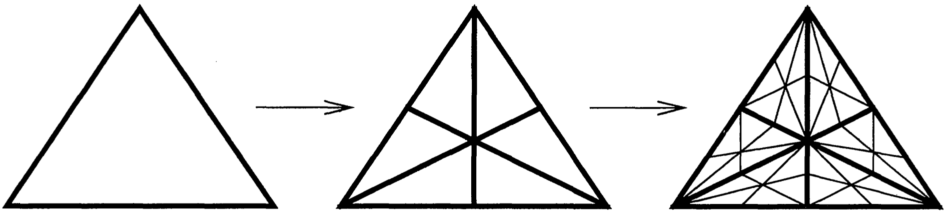


Figure 8. Barycentric subdivision in a 2-simplex, iterated twice.

labels each of the elementary simplices by owners, because each edge connects vertices of different classes.

Now the proof continues almost exactly as in the case $n = 3$: since each point in S corresponds to a cut-set, we construct a new labelling of the triangulation by asking the owner of each vertex, "Which piece would you choose if the cake were cut with these cuts?" The new auxiliary labelling is a Sperner labelling and yields nearby cut-sets that satisfy each person differently. Because this may be done with arbitrarily fine triangulations, by taking subsequences, one may find sequences of cut-sets all converging to one set of cuts in which each person chooses a different piece.

5. A CONSTRUCTIVE APPROXIMATE ALGORITHM. Notice that the preceding proof yields a constructive ϵ-approximate algorithm for cake-cutting—namely, for any prespecified ϵ (such as at the level of crumbs), one may find a set of cuts in which each person receives a piece he considers to be the best up to ϵ-tolerance in the size of the pieces. Simply start the procedure with triangulation mesh size less than ϵ, and then the "trap-door" argument gives a constructive method for finding a fully owner-labelled elementary simplex. Choosing any vertex of this simplex yields a cut-set representing the desired ϵ-approximate solution.

Such an algorithm could be implemented on a computer, which could keep track of what cuts to suggest tentatively and which player to ask, by simply following trap-doors through the simplex of cut-sets. Note that players do not have to state their preferences on every vertex in the triangulation, but only on vertices near a trap-door path, i.e., the complete auxiliary labelling may not need to be determined. So while this algorithm terminates in a number of steps bounded by the number of simplices of the triangulation, it can terminate much sooner.

We emphasize that this notion of ϵ-approximation is based on the physical size of the pieces, not on any quantitative measure of player preferences. However, if one assumes the players' measures are continuous over the simplex, then by compactness of the simplex and the finite number of players, for any $\epsilon > 0$ there exists a $\delta > 0$ such that pieces of physical size less than δ are believed by each of the players to be size less than ϵ.

6. CHORES AND RENT-PARTITIONING. Now we show how Simmons' cake-cutting method can be adapted to address other fair-division problems, such as chore division and rent-partitioning.

Finding schemes for envy-free chore division has historically been a more complicated problem than cake-cutting. Most envy-free procedures for cake-cutting do not carry over to chore division without significant modifications. Oskui

[9] solved the case for 3 people; following modifications proposed by Brams and Taylor in [2, pp. 37–39] and [3, pp. 153–55], Peterson and Su [10] gave an explicit chore division scheme for an arbitrary number of players. We now give a simpler ϵ-approximate algorithm for chore division, which falls out nicely as a special case of the *rent-partitioning* problem.

In this problem, n housemates have decided to rent an n-bedroom house for some fixed rent. Each housemate may have different preferences—one may prefer a large room, another may prefer a room with a view, etc. Is there a method for fairly dividing the rent among the rooms? We prove the following:

Rental Harmony Theorem. *Suppose n housemates in an n-bedroom house seek to decide who gets which room and for what part of the total rent. Also, suppose that the following conditions hold:*

(1) **(Good House)** *In any partition of the rent, each person finds some room acceptable.*
(2) **(Miserly Tenants)** *Each person always prefers a free room (one that costs no rent) to a non-free room.*
(3) **(Closed Preference Sets)** *A person who prefers a room for a convergent sequence of prices prefers that room at the limiting price.*

Then there exists a partition of the rent so that each person prefers a different room.

Condition (1) ensures that the problem is well-posed—one cannot talk about *preferences* if some person finds no room *acceptable*, which might happen, for instance, if the rent is too high for all rooms or the rooms are in poor condition.

The miserly condition (2) can be relaxed a bit, as we show in Section 8. The condition also rules out "free closets," i.e., rooms in which no one would live, even if free.

Condition (3) merely says that in the space of all pricing schemes, preference sets are closed in the topological sense. Note that preference sets may overlap—if in some pricing scheme a person equally prefers two rooms, that person can be assigned to either room.

The rent-partitioning problem may be viewed as a generalization of the cake-cutting problem, in which one seeks to divide *goods* fairly, and the chore division problem, in which one seeks to divide *bads* fairly. However, since the rooms (the goods) are indivisible, known cake-cutting solutions cannot be applied to this problem. And since the rental payments (the bads) are attached to specific rooms, they cannot be divided into more than n pieces and reassembled, which rules out the use of known envy-free chore-division methods such as the discrete method proposed in [3, pp. 154–55] and the procedures proposed in [10]. The two other moving knife schemes proposed for chore division in [3, pp. 153–54] guarantee each player at most $1/n$ of the chores, but are not envy-free.

Alkan, Demange, and Gale [1, pp. 1031–32] have addressed this generalization directly and offer a solution to rent-partitioning via constrained optimization. They implicitly assume conditions equivalent to our conditions (1) and (3), and use a condition weaker than condition (2), but not quite as weak as the condition (2′) that we give in Section 8.

We now show how a Sperner's lemma approach can address the rent-partitioning problem.

7. RENTAL HARMONY: CAKE-CUTTING WITH A TWIST. Our proof of the Rental Harmony Theorem follows Simmons' proof for cake-cutting, but with a twist, so we sketch it.

Suppose there are n housemates, and n rooms to assign, numbered $1, \dots, n$. Let x_i denote the price of the i-th room, and suppose that the total rent is 1. Then $x_1 + x_2 + \cdots + x_n = 1$ and $x_i \geq 0$. From this we see that the set of all pricing schemes S forms an $(n-1)$-simplex in $\mathbf{R}^n$.

Now triangulate this simplex by barycentric subdivision of small mesh size. Label it with a fully labelled vertex labelling by the names of the housemates (the same scheme as suggested for cake-cutting). The name at each vertex will be considered the "owner" of that vertex; recall that each vertex corresponds to some pricing scheme for the rooms.

Construct a new labelling from the old by asking the owner at each vertex in the triangulation: "If the rent were to be divided according to this pricing scheme, which room would you choose?" Condition (1) ensures that some answer can be given. Label the vertex by the number of the room that is answered. Let ties in preference be broken arbitrarily.

Here's the twist: the new labelling that results is quite different from the one that arose in cake-cutting. It is not a Sperner-labelling. However, because of the miserly condition (2), it has the property that along each $(n-k)$-dimensional face, k rooms are free and thus owners along that face prefer one of those k rooms. Figure 9 shows what such a labelling looks like for $n = 3$.

Is there a Sperner-like combinatorial lemma that shows the existence of a fully labelled elementary simplex in this triangulation?

If so, one could proceed as in cake-cutting, by taking finer and finer triangulations to get a sequence of fully labelled elementary simplices converging to a point,

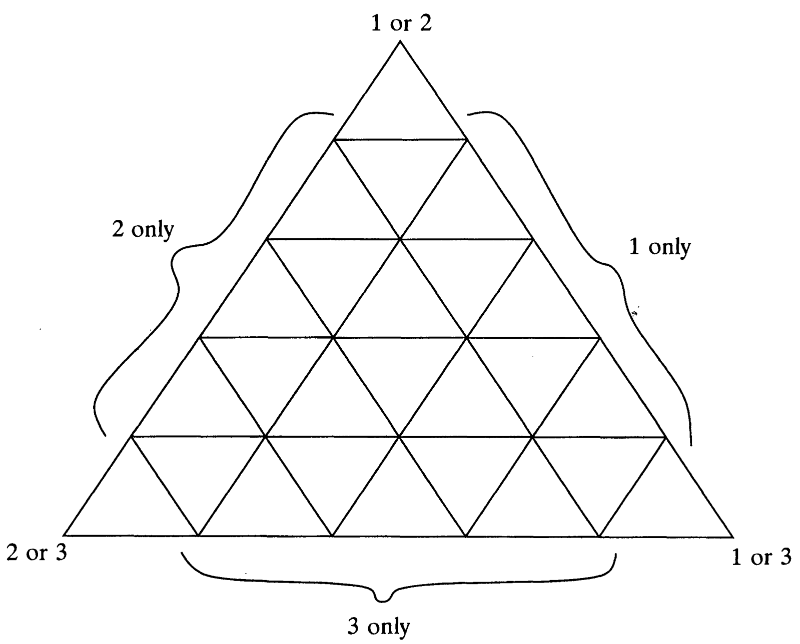


Figure 9. The dual labelling arising from rent-partitioning.

which by condition (3) yields a pricing scheme in which all housemates prefer different rooms. So, all that remains is to establish the Sperner-like combinatorial lemma with a constructive proof.

There are two ways one may proceed. The reader may enjoy proving a Sperner-like lemma for this labelling by using a trap-door argument. The interesting thing that one discovers about this labelling is that on each facet, there is only *one* fully labelled simplex that can be followed into the interior, so that the trap-door procedure succeeds without returning to the boundary again.

Or the reader may wish to prove the existence of a fully labelled simplex on the interior by appealing directly to Sperner's lemma. The key idea is to *dualize* the simplex S to form a new simplex S^*. Loosely speaking, the dual of a simplex reverses the dimensions of k-dimensional and $(n-1-k)$-dimensional faces. For instance, the corner vertices of S become the facets of S^*, and the facet barycenters of S become the vertices of S^*; see Figure 10.

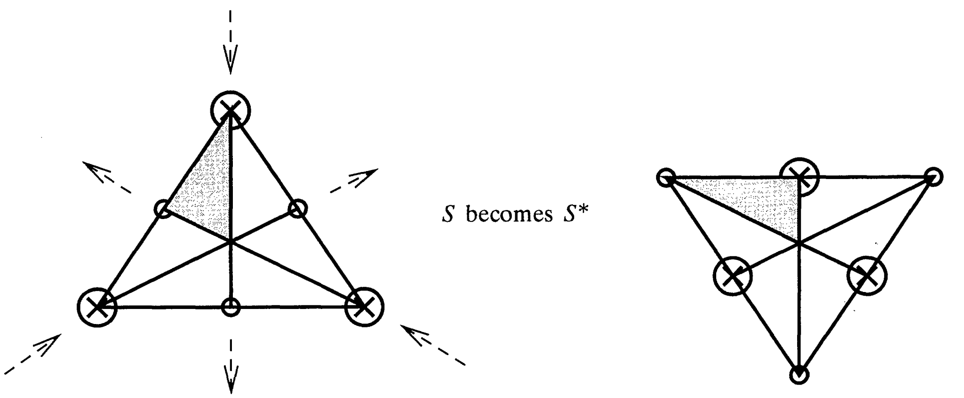


Figure 10. The dualization S^* of S. Vertices, barycenters, and one elementary simplex are marked to show how they are transformed.

A rigorous treatment of dualization can be found in Vick **[15]**. Note that S^* can be triangulated—in fact, using barycentric subdivision, the vertices and elementary simplices of S^* are in $1-1$ correspondence with the vertices and elementary simplices of S. Let the triangulation of S^* inherit a labelling via this correspondence with S. One may now verify that the labelling of S^* is a Sperner labelling! Hence there exists a fully labelled elementary simplex of S^*, which corresponds to a fully labelled elementary simplex of S, as desired. This "dual" Sperner lemma is due to Scarf **[16]**.

A constructive algorithm is obtained by following "trap-doors" in Sperner's lemma. Choose an ϵ smaller than the rental difference for which housemates wouldn't care (a penny?). Following trap-doors corresponds to suggesting pricing schemes and then asking various players, "Which piece would you choose if the rooms were priced like this?" Once a fully labelled elementary simplex is found, any point inside it corresponds to an ϵ-approximate rent-partitioning. We invite the reader to code a trap-door algorithm that could be implemented on a computer, one that would propose the necessary cut-sets and question the appropriate players at each step.

It is possible to obtain the Rental Harmony Theorem without any dualization argument and without condition (2) if one allows the possibility of negative rents. Specifically, let each person contribute a fixed amount K to a pool from which the

rent is paid. The leftover money is used to pay "rebates" associated with each room (which may be larger than K). This converts the problem into a fair division of goods (rebates), in which the space of rebates is a simplex that assumes a Sperner labelling if players demand a non-zero rebate. For large K this is quite reasonable. However, solutions may include situations in which a housemate is being *paid* by the others to live there. Thus allowing this possibility may not be realistic because in real life, paying housemates are more likely to ditch the subsidized housemate and use the extra room (and extra money) in other ways.

8. COMMENTS AND DISCUSSION. The Rental Harmony Theorem establishes the existence of envy-free chore division and a new ϵ-approximate algorithm, by simply thinking of the rent payments as chores and ignoring the rooms; divisibility of chores can be achieved by dividing the time spent on them. When reinterpreted, the three conditions from the Rental Harmony Theorem become: (1) all the chores must be assigned, (2) each person prefers no chores to some chores, and (3) preference sets are closed. These are pretty reasonable assumptions. The ϵ-approximate algorithm that arises from this does not involve a lot of cutting and reassembling, as do the exact methods proposed in **[3]** and **[10]**.

For rent-partitioning, we point out that condition (2) may not always be a reasonable assumption. For instance, someone may be willing to pay a little bit of money for a room that is slightly larger than a free room. However, by inspecting the proof, one sees that the Rental Harmony Theorem still holds with a weakened version of condition (2):

> Condition (2′). *Each person never chooses the most expensive room if there is a free room available. This does not require the person to choose the free room.*

In particular, this will hold if a person always prefers a free room to a room costing at least $1/(n-1)$ of the total rent. Hence condition (2′) is a slightly weaker sufficient condition than that given by [**1**, pp. 1031–32]. To see why the Rental Harmony Theorem still holds, consider its proof and note that using this condition gives a more complicated labelling of S, but the corresponding labelling on S^* still remains Sperner.

What condition (2′) does not address is a situation in which the total rent is so low, or some room so large, that one would be willing to pay for the most expensive room even when some other room is free. In practice, however, housemates do not usually choose a house with such lopsided arrangements. Even still, condition (2′) can likely be weakened further, but the extent to which it can (and still maintain non-negative rents) is an open question.

Other triangulations may be used instead of barycentric subdivision. These have better convergence properties but are harder to describe; see **[17]** for a survey and applications to fixed point algorithms.

9. ANECDOTE AND ACKNOWLEDGMENTS. My first exposure to the Sperner argument for cake-cutting came via Michael Starbird, who attributed the method to a graduate student of his, Forest Simmons. Simmons had been presenting this cake-cutting scheme to math clubs and high school groups, but never formally submitted the idea for publication. His inspiration was the MONTHLY article by Stromquist **[14]**, which made use of a theorem that can be proved by Sperner's lemma.

Many years later, when my friend Brad Mann told me about the rent-partitioning dilemma that he and his housemates were facing, I was reminded of these ideas and realized that Sperner's lemma could also be adapted to treat rent-partitioning, as well as chore division.

I am grateful to Arthur Benjamin, Steven Brams, Brad Mann, Forest Simmons, and Ravi Vakil for many helpful discussions, and I thank Michael Starbird for introducing me to Sperner's lemma.

REFERENCES

1. A. Alkan, G. Demange, and D. Gale, Fair allocation of indivisible goods and criteria of justice, *Econometrica* **59** (1991) 1023–1039.
2. S. J. Brams and A. D. Taylor, *An envy-free cake division algorithm*, Economic Research Reports, C. V. Starr Center for Applied Economics, New York University, 1992.
3. S. J. Brams and A. D. Taylor, *Fair Division: from Cake-Cutting to Dispute Resolution*, Cambridge University Press, Cambridge, 1996.
4. S. J. Brams and A. D. Taylor, An envy-free cake division protocol, *Amer. Math. Monthly* **102** (1995) 9–18.
5. D. I. A. Cohen, On the Sperner lemma, *J. Combin. Theory* **2** (1967) 585–587.
6. M. Gardner, *aha! Insight*. W. F. Freeman and Co., New York, 1978.
7. B. Knaster, C. Kuratowski, and S. Mazurkiewicz, Ein Beweis des Fixpunktsatzes für n-dimensionale Simplexe, *Fund. Math.* **14** (1929) 132–137.
8. H. W. Kuhn, Simplicial Approximation of Fixed Points, *Proc. Nat. Acad. Sci. U.S.A.* **61** (1968) 1238–1242.
9. R. Oskui, Dirty work problem, preprint.
10. E. Peterson and F. E. Su, Exact procedures for envy-free chore division, preprint.
11. J. M. Robertson and W. A. Webb, *Cake-Cutting Algorithms: Be Fair If You Can*, A K Peters Ltd., Natick, Mass., 1998.
12. F. W. Simmons, private communication to Michael Starbird, 1980.
13. E. Sperner, Neuer Beweis für die Invarianz der Dimensionszahl und des Gebietes, *Abh. Math. Sem. Univ. Hamburg* **6** (1928) 265–272.
14. W. Stromquist, How to cut a cake fairly, *Amer. Math. Monthly* **87** (1980) 640–644.
15. J. W. Vick, *Homology Theory*. Springer-Verlag, New York, 1994.
16. H. Scarf, *The Computation of Equilibrium Prices: An Exposition*, Cowles Foundation Discussion Paper No. 473, Cowles Foundation for Research in Economics at Yale University, November, 1977.
17. M. J. Todd, *The Computation of Fixed Points and Applications*, Lecture Notes in Economic and Mathematical Systems, New York, 1976.

FRANCIS SU earned his Ph.D. from Harvard in 1995 studying random walks, but his interest in fair division was a result of thesis procrastination. His research is now (fairly? randomly?) divided between studies of random walk convergence and fair division algorithms. He is an assistant professor at Harvey Mudd College, a "blue dot" MAA Project NExT Fellow, and an amateur songwriter in his spare time. Since the writing of this article he and HMC undergraduate Elisha Peterson wrote a web applet that uses an improved trapdoor algorithm; their *Fair Division Calculator* works for cakes, chores, or rent and they invite you to try it out at `http://www.math.hmc.edu/~su/fairdivision/`.
Harvey Mudd College, Claremont, CA 91711
su@math.hmc.edu

NOTES

Edited by **Jimmie D. Lawson and William Adkins**

Which Tanks Empty Faster?

Leonid G. Hanin

Suppose that water towers like those shown in Figure 1 are initially filled with water and have the same volume, height, and cross-sectional outlet area. Which one empties first? This problem arises naturally when designing water-supplying tanks or funnels.

We find a formula expressing the emptying time as a function of the volume of liquid and its initial height. We compute the emptying time for several specific tank shapes, in particular, for those shown in Figure 1. We also address the question whether there exists a tank with a given volume and height for which the emptying time is minimal.

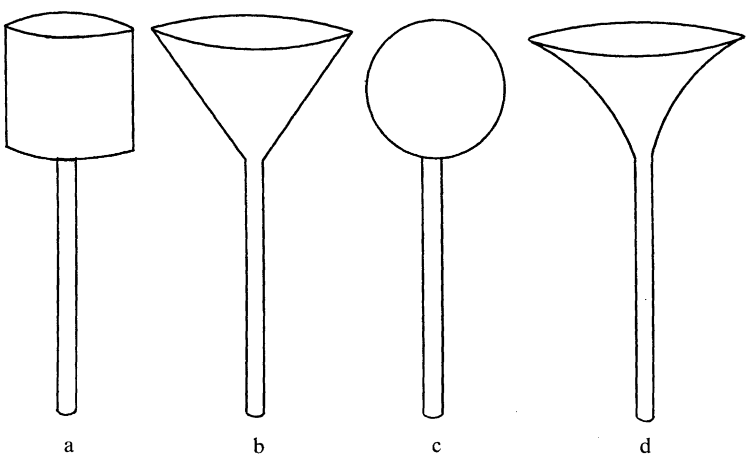


Figure 1

1. Mathematical Model. Suppose a tank of volume V and height H is initially filled with an incompressible liquid. A small (but not microscopic) hole with cross-sectional area S is made in the bottom of the tank. Let $A(h)$, $0 \leq h \leq H$, be the area of the tank cross-section at height h. We assume that the function $A(h)$ is continuous.

Let h be the height of the liquid in the tank at a time t. Let Δh be the height drop during a small amount of time Δt that elapses from the moment t. Then the volume decrease ΔV approximately equals $A(h)\Delta h$. As we show later, the velocity of the outgoing flow is a function of liquid height: $v = v(h)$. Hence, the volume of

the liquid leaving the tank during the time Δt is approximately equal to $Sv(h)\Delta t$. Thus, $A(h)\Delta h \simeq -Sv(h)\Delta t$. Letting $\Delta t \to 0$, we obtain the differential equation

$$h' = -\frac{S}{A(h)}v(h). \tag{1}$$

To solve it, the function $v(h)$ must be specified.

2. Torricelli's Law. In 1640, E. Torricelli found that

$$v(h) = \sqrt{2gh}, \tag{2}$$

where g is the acceleration due to gravity. Here is a simple argument for (2); see also [**1**] and [**2**]. Let Δm be the mass of the liquid leaving the tank during the time Δt. Then the potential energy loss $\Delta\Pi$ is approximately equal to Δmgh. The kinetic energy ΔK of the equal amount of liquid flowing out of the tank through the orifice during the time Δt is about $\Delta mv^2(h)/2$. Equating $\Delta\Pi$ and ΔK gives (2). For a careful derivation of Torricelli's formula, see [**3**, pp. 47–48 and 56–59].

In reality, due to viscosity of the liquid, its rotation, and constriction of the jet emerging from the tank, (2) is not quite accurate, especially in the case of non-horizontal outflow. Experiments show that for a circular orifice

$$v(h) = \alpha\sqrt{gh}, \tag{3}$$

where the constant α depends on the physical properties of the liquid [**3**, pp. 47–48]. For example, the approximate value of the coefficient α for water is 0.84.

If liquid were oozing from the tank at the constant initial rate $v_0 = v(H) = \alpha\sqrt{gH}$, then the emptying time T^* would be

$$T^* = \frac{V}{Sv_0} = \frac{V}{S\alpha\sqrt{gH}}. \tag{4}$$

However, according to Torricelli's law, the efflux rate is decreasing with the decrease of height. Therefore, the emptying time T is

$$T = kT^* = k\frac{V}{S\alpha\sqrt{gH}}, \tag{5}$$

where $k > 1$. In general, the coefficient k depends on the height H and the shape of the tank. We show, however, that for many practically important tank forms, the coefficient k is an absolute constant.

3. Emptying Time. With (3) taken into account, the differential equation (1) takes on the following form:

$$h' = -S\alpha\sqrt{g}\,\frac{\sqrt{h}}{A(h)}. \tag{6}$$

Some properties of this equation are discussed in [**1**] and [**2**]. A classroom demonstration based on this equation for a cylindrical container is described in [**4**].

The solution $h = h(t)$ of (6) satisfying the initial condition $h(0) = H$ is given implicitly by

$$\int_h^H \frac{A(u)}{\sqrt{u}}\,du = S\alpha\sqrt{g}\,t. \tag{7}$$

Since $h(T) = 0$, we obtain from (7) that

$$T = \frac{1}{S\alpha\sqrt{g}}\int_0^H \frac{A(u)}{\sqrt{u}}\,du. \tag{8}$$

This formula provides a closed expression for the emptying time T.

Observing that

$$V = \int_0^H A(u)\, du \tag{9}$$

we rewrite (8) in the form (5), where

$$k = k(H) = \sqrt{H}\, \frac{\int_0^H \left(A(u)/\sqrt{u} \right) du}{\int_0^H A(u)\, du}. \tag{10}$$

We set

$$g(s) := A(s^2), \quad 0 \le s \le \sqrt{H}, \tag{11}$$

in (9) and (10) to find that

$$V = 2\int_0^{\sqrt{H}} g(s) s\, ds \quad \text{and} \quad \int_0^H \frac{A(u)}{\sqrt{u}}\, du = \int_0^{\sqrt{H}} g(s)\, ds. \tag{12}$$

This leads to the following alternative expressions for k:

$$k = \sqrt{H}\, \frac{\int_0^{\sqrt{H}} g(s)\, ds}{\int_0^{\sqrt{H}} g(s) s\, ds} = \frac{\int_0^1 g(\sqrt{H} s)\, ds}{\int_0^1 g(\sqrt{H} s) s\, ds}. \tag{13}$$

The case of a circularly symmetric tank is probably the most important. The lateral surface of such tank is obtained by rotating the graph of a nonnegative continuous function $f(h)$, $0 \le h \le H$, about the h axis. Then

$$A(h) = \pi f^2(h). \tag{14}$$

Suppose f is homogeneous of some order $\theta \ge 0$: that is, for any $\lambda > 0$ and for all admissible $h \in [0, H]$,

$$f(\lambda h) = \lambda^\theta f(h). \tag{15}$$

Then the function g defined by (11) is homogeneous of order 4θ. In view of (13), this leads to the important conclusion that in this case the coefficient k depends only on f, that is, only upon the shape of the tank.

We compute the coefficient k for a few simple and widely used tank shapes, including those in Figure 1. Formula (5) then gives the emptying time.

Cylinder. Let the tank be a right circular cylinder of height H with base radius R, where $R^2 = V/(\pi H)$; see Figure 1a. In this case, $f(h) = R$, $0 \le h \le H$, which is a homogeneous function of order 0. Then $g(s) = \pi R^2$, and therefore by (13), $k = 2$.

Cone. For the tank in the form of a right circular cone (Figure 1b) with height H and radius R, where $R^2 = 3V/(\pi H)$, we have $f(h) = \gamma h$ with $\gamma = H/R$. Then f satisfies (15) with $\theta = 1$, and it follows easily from (13) that $k = 1.2$.

Frustum of a cone. Suppose the tank has the form of a right circular frustum of a cone with lower base radius R_1 and upper base radius R_2. Then $f(h) = a + bh$, where $a = R_1$ and $b = (R_2 - R_1)/H$, and $g(s) = \pi(a + bs^2)^2$. A straightforward calculation based on (13) gives the following expression for the coefficient k:

$$k = \frac{2}{5} \cdot \frac{8R_1^2 + 4R_1R_2 + 3R_2^2}{R_1^2 + R_1R_2 + R_2^2}. \tag{16}$$

Thus, k is independent of H. For $R_1 = 0$, (16) yields the value 1.2 already obtained for the cone. In the other extreme case of the inverse cone ($R_2 = 0$), we have $k = 3.2$. For $R_1 = R_2$, (16) produces the value $k = 2$ already found earlier for the cylinder.

Spherical tanks. Let the tank be a truncated sphere of height H, which is the most popular form for aquariums. The radius R of the sphere is determined by the tank volume through the formula $V = \pi H^2(R - H/3)$. In this case, $f^2(h) = h(2R - h)$, $0 \le h \le H$. Hence, $g(s) = \pi s^2(2R - s^2)$, and by (13) we obtain after a short calculation that

$$k(H) = \frac{2}{5} \cdot \frac{10R - 3H}{3R - H}.$$

In particular, for a hemispherical tank ($H = R$), we find that $k = 1.4$ while for a complete spherical tank ($H = 2R$; see Figure 1c), we obtain $k = 1.6$.

Table 1 summarizes our results and shows the relative emptying efficiency of various tank forms. The conic shape turns out to be significantly more efficient than other natural shapes. This explains why it is so widely used for funnels. Formula (5) and Table 1 allow us to compare emptying times of tanks of various shapes with variable volume and height.

TABLE 1 "Emptying efficiencies" of different tank shapes

Tank Shape	Cone	Hemisphere	Sphere	Cylinder	Inverse cone
k	1.2	1.4	1.6	2	3.2

For physical reasons, the coefficient k is always larger than 1. Can it be less than 1.2? As shown in the next section, the answer to this question is YES!

4. Are there Tanks with the Minimal Emptying Time? Let the function that determines the shape of a circularly symmetric tank be

$$f(h) = Ch^{\mu}, \quad 0 \le h \le H, \tag{17}$$

with some constants $\mu \ge 0$ and $C > 0$. Given μ, the value of C can be found from the relation $V = \pi C^2 H^{2\mu+1}/(2\mu + 1)$. The function (17) is homogeneous of order $\theta = \mu$. Then $g(s) = \pi C^2 s^{4\mu}$, and using (13) we obtain easily that

$$k = \frac{4\mu + 2}{4\mu + 1}. \tag{18}$$

For $\mu = 0$ and $\mu = 1$, we recover the values of $k = 2$ for the cylinder and $k = 1.2$ for the cone, respectively. For $\mu = 2$, we have $k = 10/9$, which means that, for the parabolic tank shown in Figure 1d, the emptying time is more than 7% smaller than for the corresponding conic tank in Figure 1b.

It follows from (18) that, for power functions (17) with large μ, the coefficient k can be as close to 1 as we wish. Therefore, the emptying time can be arbitrarily close to its theoretical minimum (4). We show, however, that the minimum is not attained. This means that among all tanks with a given volume and height none has the minimal emptying time. Our argument also provides a mathematical proof that $k > 1$.

Consider tanks with a given volume V and height H. We continue to assume that the cross-sectional area $A(h)$ at height h is a continuous function of h.

According to (8) and (12), we are dealing with the extremal problem

$$\mathscr{T}(g) := \int_0^a g(s)\,ds \to \min$$

subject to the constraint $\int_0^a g(s) s\,ds = b$, where $a = \sqrt{H}$, $b = V$, and g belongs to the class G of nonnegative continuous functions on $[0, a]$. Consider a similar problem

$$\mathscr{T}(\nu) := \nu([0,a]) \to \min, \quad \int_0^a s\,d\nu(s) = b \tag{19}$$

on the larger class N of nonnegative finite Borel measures ν on $[0, a]$. For every $\nu \in N$, we have

$$\mathscr{T}(\nu) = \int_0^a d\nu(s) \geq \frac{1}{a}\int_0^a s\,d\nu(s) = \frac{b}{a} = \mathscr{T}\left(\frac{b}{a}\delta_a\right),$$

where δ_a is the Dirac measure at a. Therefore, the measure $\nu^* := b\delta_a/a$ is a minimizer of the functional $\mathscr{T}$ on the set N, and the minimum value of $\mathscr{T}$ on N is equal to b/a. Taking a clue from (13), we find that the corresponding minimal value of k is

$$k^* = a\frac{\int_0^a d\nu^*(s)}{\int_0^a s\,d\nu^*(s)} = 1.$$

If for some $\nu \in N$ we have $\mathscr{T}(\nu) = b/a$, then $\int_0^a (a-s)\,d\nu(s) = 0$, whence it follows that ν is proportional to the Dirac measure at a. Therefore, (19) ensures that $\nu = \nu^*$. Thus, the minimizer ν^* is unique. This implies that the infimum of the functional $\mathscr{T}$ on the set G is not attained. However, there are sequences of functions in G for which the corresponding values of the functional $\mathscr{T}$ converge to b/a. One of them is a sequence of functions g_n that are related via (11) and (14) to functions (17) with a sequence μ_n tending to infinity.

ACKNOWLEDGMENTS. I thank Carol and Robert Fisher for their many suggestions on stylistic improvement of the manuscript, and my sons Mark and Boris for producing the figure and table.

REFERENCES

1. R. D. Driver, Torricelli's law—an ideal example of an elementary ODE, *Amer. Math. Monthly* **105** (1998) 453–455.
2. C. W. Groetch, Inverse problems and Torricelli's law, *College Math. J.* **24** (1993) 210–217.
3. W. Kaufmann, *Fluid Mechanics*, McGraw-Hill Book Company, New York, 1963.
4. T. Farmer, and F. Gass, Physical demonstrations in the calculus classroom, *College Math. J.* **23** (1992) 146–148.

Idaho State University, Pocatello, ID 83209-8085
hanin@isu.edu

Two Uniformly Distributed Parameters Defining Catalan Numbers

David Callan

A *path* is a finite sequence of ± 1's with a graphical representation as a sequence of contiguous steps of slope $+1$ (upsteps) and -1 (downsteps). For example, the path $w = (1, -1, -1, 1, -1, 1, 1, -1)$ is pictured in Figure 1.

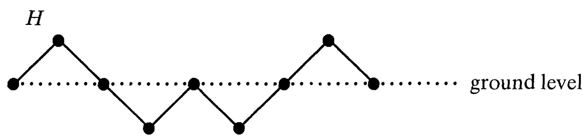


Figure 1

Let $\mathscr{P}_n$ denote the set of $\binom{2n}{n}$ paths consisting of n upsteps and n downsteps. Each path in $\mathscr{P}_n$ starts and terminates at "ground level" as in Figure 1. There is a well known parameter (statistic) on $\mathscr{P}_n$ that we will call *northcnt* (to suggest a count north of a baseline). For $w \in \mathscr{P}_n$, northcnt(w) is the number of w's n upsteps that lie above ground level. Thus northcnt = 2 in Figure 1, and as w ranges over $\mathscr{P}_n$ northcnt has possible values 0 through n. The paths for which northcnt = n—that is, the paths that lie entirely at or above ground level—we call *Catalan* paths. Dually, we call the paths with northcnt = 0 *inverted Catalan* paths: reflection in ground level gives a bijection between the two classes. It is a famous fact that exactly $1/(n+1)$ of the paths in $\mathscr{P}_n$ are Catalan: they are counted by the Catalan number $\frac{1}{n+1}\binom{2n}{n}$. A combinatorially satisfying way to see this is via the Chung-Feller Theorem, which asserts that the parameter northcnt is in fact *uniformly* distributed on $[0, n]$. This partitions $\mathscr{P}_n$ into $n+1$ equal-size classes, one of which consists of the Catalan paths. For combinatorial proofs of the Chung-Feller Theorem, see **[1]**, **[2]**, **[3]**, or **[4]**.

Curiously, there is another parameter on $\mathscr{P}_n$, *westcnt*, that serves the same purpose: it is also uniformly distributed on $[0, n]$ and it has a constant value on the set of inverted Catalan paths. To define westcnt(w), let H denote the highest point of w, taking the leftmost one if there is more than one highest point as in Figure 1. Then westcnt(w) is the number of w's n upsteps that lie to the left (west) of H. Thus the path in Figure 1 has westcnt = 1, and westcnt = 0 precisely for the inverted Catalan paths. The parameter westcnt is implicit in **[5]**.

One could show directly that westcnt is uniformly distributed on $[0, n]$. This is essentially done in **[5]**, modulo translation from bracket sequences to lattice paths. But that still leaves open the question, why? Can one "explain" why northcnt and westcnt are equidistributed? A satisfactory answer would consist of a "nice" bijection $\phi : \mathscr{P}_n \to \mathscr{P}_n$ such that westcnt(w) = northcnt($\phi(w)$) for all $w \in \mathscr{P}_n$. Here we give a simple such bijection.

To define ϕ, first observe that every path in $\mathscr{P}_n$ can be uniquely decomposed as in Figure 2 where the C_i and D_i are inverted Catalan paths (possibly empty), lying

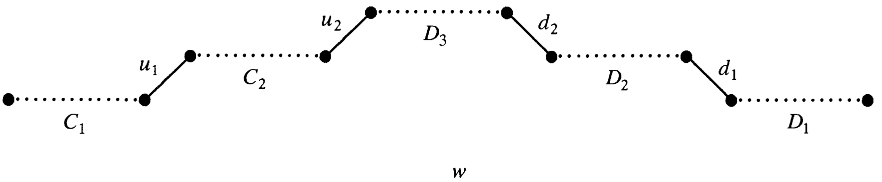


Figure 2

below the dotted segments. Each u_i is an upstep and each d_i is a downstep. There will be k C's and $k+1$ D's for some $k \geq 0$; in the illustration, $k = 2$. To see uniqueness, imagine the space above ground level divided into horizontal strips as indicated by the dotted lines (extended) in Figure 2. Then u_i, d_i are respectively the leftmost upstep and rightmost downstep in the ith strip above ground level.

The path $\phi(w)$ is given by flipping over each C_i path so it becomes a Catalan path C_i' and then rearranging components as in Figure 3. Note that since H (the

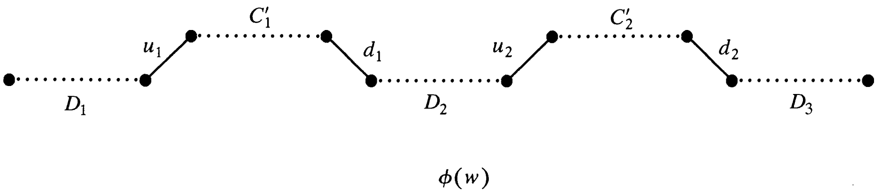


Figure 3

leftmost high point) is the northeast tip of of u_k (of u_2 in Figure 2)

$$\text{westcnt}(w) = \#\, u\text{'s} + \text{total}\ \#\ \text{upsteps in the}\ C_i.$$

Also,

$$\text{northcnt}(\phi(w)) = \#\, u\text{'s} + \text{total}\ \#\ \text{upsteps in the}\ C_i'.$$

However, for each i, # upsteps in C_i = # downsteps in C_i = # upsteps in C_i', and hence westcnt(w) = northcnt($\phi(w)$), as desired.

Finally, to show ϕ is a bijection, we must check reversibility: can the u_i, d_i, C_i', D_i as in Figure 3 be retrieved uniquely from each path in $\mathscr{P}_n$? Yes: consider the first horizontal strip above ground level. Traversing this strip left to right, upsteps and downsteps are encountered alternately. These determine the u_i and d_i (if any). The connecting paths (possibly empty) determine the C_i' and D_i in order. We are done.

REFERENCES

1. M. D. Atkinson and J.-R. Sack, Generating binary trees at random, *Inform. Process Lett.* **41** (1992) 21–23.
2. D. Callan, Pair them up!: A visual approach to the Chung-Feller theorem, *College Math. J.* **26** (1995) 196–198.
3. H. M. Finucan, *Proc. Fourth Australian Conf., Univ. Adelaide, Adelaide, Some elementary aspects of the Catalan numbers, Lecture Notes in Math., Vol. 560*, Springer, Berlin 1976, 41–45.
4. T. V. Narayana, Cyclic permutation of lattice paths and the Chung-Feller theorem, *Skandinavisk Aktuarietidskrift* **50** (1967) 23–30.
5. D. Rubinstein, Catalan numbers revisited, *J. Combin. Theory Ser. A* **68** (1994) 486–490.

Department of Statistics, University of Wisconsin-Madison, 1210 W. Dayton Street, Madison, WI 53706-1693
callan@stat.wisc.edu

Determinants of Commuting-Block Matrices

Istvan Kovacs, Daniel S. Silver, and Susan G. Williams

Let $\mathscr{R}$ be a commutative ring, and let $\mathrm{Mat}_n(\mathscr{R})$ denote the ring of $n \times n$ matrices over $\mathscr{R}$. We can regard a $k \times k$ matrix $M = (A^{(i,j)})$ over $\mathrm{Mat}_n(\mathscr{R})$ as a *block matrix*, a matrix that has been partitioned into k^2 submatrices (*blocks*) over $\mathscr{R}$, each of size $n \times n$. When M is regarded in this way, we denote its determinant in $\mathscr{R}$ by $|M|$. We use the symbol $D(M)$ for the determinant of M viewed as a $k \times k$ matrix over $\mathrm{Mat}_n(\mathscr{R})$. It is important to realize that $D(M)$ is an $n \times n$ matrix.

Theorem 1. *Let $\mathscr{R}$ be a commutative ring. Assume that M is a $k \times k$ block matrix of blocks $A^{(i,j)} \in \mathrm{Mat}_n(\mathscr{R})$ that commute pairwise. Then*

$$|M| = |D(M)| = \left| \sum_{\pi \in S_k} (\operatorname{sgn} \pi) A^{(1,\pi(1))} A^{(2,\pi(2))} \cdots A^{(k,\pi(k))} \right|. \tag{1}$$

Here S_k is the symmetric group on k symbols; the summation is the usual one that appears in the definition of determinant. Theorem 1 is well known in the case $k = 2$; the proof is often left as an exercise in linear algebra texts; see [**4**, p. 164]. The general result is implicit in [**3**], but it is not widely known. We present a short, elementary proof using mathematical induction on k. We sketch a second proof when the ring $\mathscr{R}$ has no zero divisors, a proof that is based on [**3**] and avoids induction by using the fact that commuting matrices over an algebraically closed field can be simultaneously triangularized.

Proof: We use induction on k. The case $k = 1$ is evident. We suppose that (1) is true for $k - 1$ and then prove it for k. Observe that the following matrix equation holds:

$$\begin{pmatrix} I & O & \cdots & O \\ -A^{(2,1)} & I & \cdots & O \\ \vdots & \vdots & \cdots & \vdots \\ -A^{(k,1)} & O & \cdots & I \end{pmatrix} \begin{pmatrix} I & O & \cdots & O \\ O & A^{(1,1)} & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & A^{(1,1)} \end{pmatrix} M = \begin{pmatrix} A^{(1,1)} & * \; * \; * \\ O & \\ \vdots & N \\ O & \end{pmatrix},$$

where N is a $(k-1) \times (k-1)$ matrix. For the sake of notation, we write this as

$$PQM = R, \tag{2}$$

where the symbols are defined appropriately. By the multiplicative property of determinants we have $D(PQM) = D(P)D(Q)D(M) = (A^{(1,1)})^{k-1} D(M)$ and $D(R) = A^{(1,1)} D(N)$. Hence we have $(A^{(1,1)})^{k-1} D(M) = A^{(1,1)} \; D(N)$. Take the determinant of both sides of the last equation. Using $|D(N)| = |N|$, a consequence of the induction hypothesis, together with (2), we find

$$\begin{aligned} |A^{(1,1)}|^{k-1} |D(M)| &= |A^{(1,1)}| \, |D(N)| = |A^{(1,1)}| \, |N| \\ &= |R| = |P| \, |Q| \, |M| = |A^{(1,1)}|^{k-1} |M|. \end{aligned}$$

If $|A^{(1,1)}|$ is neither zero nor a zero divisor, then we can divide the sides by $|A^{(1,1)}|^{k-1}$ to get (1). For the general case, we embed $\mathscr{R}$ in the polynomial ring

$\mathscr{R}[z]$, where z is an indeterminant, and replace $A^{(1,1)}$ by the matrix $zI + A^{(1,1)}$. Since the determinant of $zI + A^{(1,1)}$ is a monic polynomial of degree n, and hence is neither zero nor a zero divisor, (1) holds again. Substituting $z = 0$ (equivalently, equating constant terms of both sides) yields the desired result. ∎

We sketch an alternative proof of Theorem 1 when $\mathscr{R}$ has no zero divisors, a proof suggested to us by the referee. It is based on ideas of [**3**]; see also [**1**]. If $\mathscr{R}$ is a commutative ring with no zero divisors, then we can embed it in its quotient field and then pass to the algebraic closure F. We now regard the blocks $A^{(i,j)}$ as operators on the vector space F^n, and M as an operator on the tensor product $V = F^n \otimes F^k$. Since the blocks $A^{(i,j)}$ commute pairwise, there exists a basis for F^n with respect to which each $A^{(i,j)}$ is upper triangular; see [**2**, p. 108]. We form the tensor product of such a basis with the standard one for F^k, thereby constructing a new basis for V. The change of basis has the effect on M of simultaneously triangularizing each block. Thus it suffices to assume that each block $A^{(i,j)}$ is upper triangular.

The matrix M is permutation-similar to a $n \times n$ block matrix $\tilde{M} = (\tilde{A}_{p,q})$ such that $\tilde{A}_{p,q} = (A^{(i,j)}_{p,q})$ is an $k \times k$ matrix consisting of the p, q-entries of the $A^{(i,j)}$. Since each $A^{(i,j)}$ is upper triangular, $\tilde{A}_{p,q} = 0$ whenever $p > q$. Hence $|\tilde{M}| = |\tilde{A}_{1,1}| \cdots |\tilde{A}_{n,n}| = \prod_{r=1}^{n} \sum_{\pi \in S_k} (\operatorname{sgn} \pi) A^{(1,\pi(1))}_{r,r} \cdots A^{(k,\pi(k))}_{r,r}$. Since each $A^{(i,j)}$ is upper triangular, the last product is equal to $\prod_{r=1}^{n} \sum_{\pi \in S_k} (\operatorname{sgn} \pi)(A^{(1,\pi(1))} \cdots A^{(k,\pi(k))})_{r,r}$. But this is equal to $|\sum_{\pi \in S_k} (\operatorname{sgn} \pi) A^{(1,\pi(1))} \cdots A^{(k,\pi(k))}|$. Hence (1) holds.

The second proof shows that the commutativity hypotheses of Theorem 1 can be replaced by the weaker condition that the blocks can be simultaneously triangularized. However, *some* hypothesis about the blocks is certainly needed for the conclusion of the theorem to hold, as the reader can see by considering the matrix

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

We conclude by describing a class of block matrices that satisfy the commutativity hypothesis of Theorem 1. Matrices of this type arose in [**5**], and were the original motivation for this investigation. Let $p^{(i,j)}(t)$ be polynomials, $1 \le i,j \le k$, and let N be an $n \times n$ matrix. All coefficients are in $\mathscr{R}$, which can be taken to be the field of complex numbers, if the reader desires. Since the matrices $p^{(i,j)}(N)$ commute pairwise, the block matrix

$$M = \begin{pmatrix} p^{(1,1)}(N) & \cdots & p^{(1,k)}(N) \\ \vdots & \ddots & \vdots \\ p^{(k,1)}(N) & \cdots & p^{(k,k)}(N) \end{pmatrix}$$

satisfies the hypothesis of Theorem 1. In fact, using the theorem we can say something about the determinant of M. Let $p(t)$ be the determinant of

$$\begin{pmatrix} p^{(1,1)}(t) & \cdots & p^{(1,k)}(t) \\ \vdots & \ddots & \vdots \\ p^{(k,1)}(t) & \cdots & p^{(k,k)}(t) \end{pmatrix},$$

and let $\zeta_1, \dots, \zeta_n$ be the (not necessarily distinct) eigenvalues of N. We leave the proof of the following assertion as an exercise:

$$|M| = \prod_{r=1}^{n} p(\zeta_r).$$

REFERENCES

1. R. Bhatia, R. A. Horn, and F. Kittaneh, Normal approximants to binormal operators, *Linear Alg. Appl.* **147** (1991), 169–179.
2. R. R. Halmos, *Finite-dimensional Vector Spaces*, Springer-Verlag, New York, 1993.
3. R. A. Horn, Solution to Problem 96–11, *SIAM Review* **39** (1997), 518–519.
4. K. Hoffman and R. Kunze, *Linear Algebra*, Prentice-Hall, New Jersey, 1971.
5. D. S. Silver and S. G. Williams, Coloring link diagrams with a continuous palette, *Topology*, to appear.

Department of Mathematics and Statistics, University of South Alabama, Mobile, AL 36688
silver@mathstat.usouthal.edu, williams@mathstat.usouthal.edu

Mixtilinear Incircles

Paul Yiu

L. Bankoff [**1**] has coined the term *mixtilinear incircles* of a triangle for the three circles each tangent to two sides and to the circumcircle internally. Consider a triangle ABC and its mixtilinear incircle in the angle A, with center K_A, and radius ρ_A. Bankoff has established the fundamental formula

$$r = \rho_A \cdot \cos^2 \frac{\alpha}{2}, \tag{1}$$

where r is the inradius of the triangle, and α is the magnitude of the angle at A. This formula had appeared earlier as an exercise in [**2**, p. 23]. It leads to a simple construction of the mixtilinear incircle. Denote by I the incenter of triangle ABC, and let the perpendicular through I to the bisector of angle A intersect the sides AC, AB at Y_1 and Z_1, respectively. The perpendiculars at these points to their respective sides intersect again on the angle bisector, at the mixtilinear incenter K_A. The circle with center K_A, passing through Y_1 (and Z_1), is the mixtilinear incircle in angle A; see Figure 1.

In this note, we demonstrate the usefulness of the notion of barycentric coordinates in discovering remarkable geometric properties relating to the mixtilinear incircles of a triangle. To keep the note self-contained, we refrain from using (1), except for the remarks at the end.

Denote by A' the point of contact of the mixtilinear incircle in angle A with the circumcircle. For convenience, we denote K_A by K, and ρ_A by ρ when there is no danger of confusion; see Figure 2. The center K lies on the bisector of angle A, and $AK : KI = \rho : -(\rho - r)$. In terms of barycentric coordinates,

$$K = \frac{1}{r}[-(\rho - r)A + \rho I]. \tag{2}$$

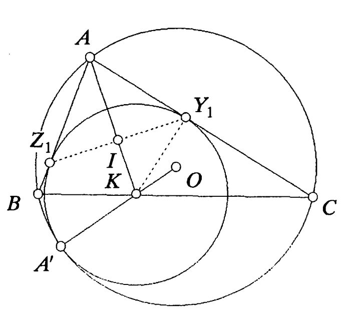


Figure 1

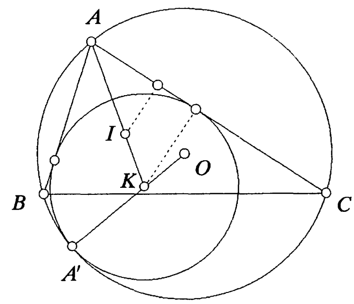


Figure 2

Also, since the circumcircle $O(A')$ and the mixtilinear incircle $K(A')$ touch each other at A', we have $OK : KA' = R - \rho : \rho$, where R is the circumradius. From this,

$$K = \frac{1}{R}[\rho O + (R - \rho)A']. \tag{3}$$

Comparing (2) and (3), we obtain, by rearranging terms,

$$\frac{RI - rO}{R - r} = \frac{R(\rho - r)A + r(R - \rho)A'}{\rho(R - r)}. \tag{4}$$

We note some interesting consequences of this formula. First of all, it gives the intersection of the lines joining AA' and OI. Note that the point P on the line OI represented by the left hand side of (4) is the external center of similitude of the circumcircle and the incircle of the given triangle. This, by definition, is the point dividing the segment OI externally in the ratio of the radii of the circles. As such, it can be constructed as the intersection of the lines OI and MD, where M is the intersection of the bisector of angle A with the circumcircle, and D the point of contact of the incircle with the side BC; see Figure 3.

The same reasoning applied to the other two mixtilinear incircles shows that each of the lines AA', BB', CC' passes through the same point P on the line OI; see Figure 4.

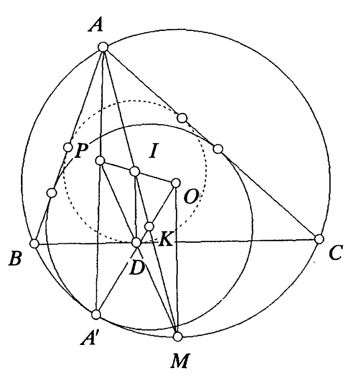


Figure 3

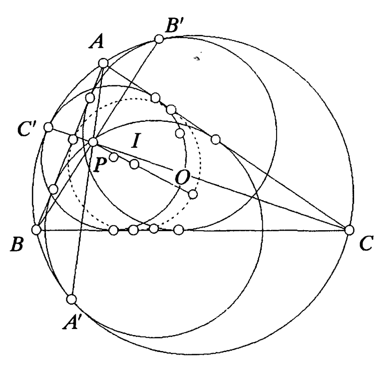


Figure 4

Theorem 1. *The three lines each joining a vertex to the point of contact of the circumcircle with the mixtilinear incircle in the angle of the vertex are concurrent at the external center of similitude of the circumcircle and the incircle.*

Equation (4) also leads to an alternative construction of the mixtilinear incircle, without the use of (1).

Construction 2. Given a triangle ABC, let P be the external center of similitude of the circumcircle (O) and incircle (I). Extend AP to intersect the circumcircle at A'. The intersection of AI and $A'O$ is the center K_A of the mixtilinear incircle in angle A.

Theorem 1 means that the triangles ABC and $A'B'C'$ are in perspective. By Desargues' Theorem, the intersections of the three pairs of lines BC, $B'C'$; CA, $C'A'$, and AB, $A'B'$ are collinear. The intersection X of the lines BC and $B'C'$ is indeed the external center of similitude of the mixtilinear incircles (K_B) and (K_C). This is clear from the following lemma, whose proof we omit.

Lemma 3. *If two distinct circles are tangent to a third circle, both internally or both externally, then the line joining the points of contact passes through the external center of similitude of the two circles.*

If one of the tangencies is internal and the other is external, then the line joining the points of contact passes through the internal center of similitude of the two circles; see Figure 5.

It is easy to determine the barycentric coordinates of X with respect to B and C. In fact,

$$X = \frac{\rho_C \cdot K_B - \rho_B \cdot K_C}{\rho_C - \rho_B} = \frac{-\left(1 - \frac{r}{\rho_B}\right)B + \left(1 - \frac{r}{\rho_C}\right)C}{\left(\frac{1}{\rho_B} - \frac{1}{\rho_C}\right)r}.$$

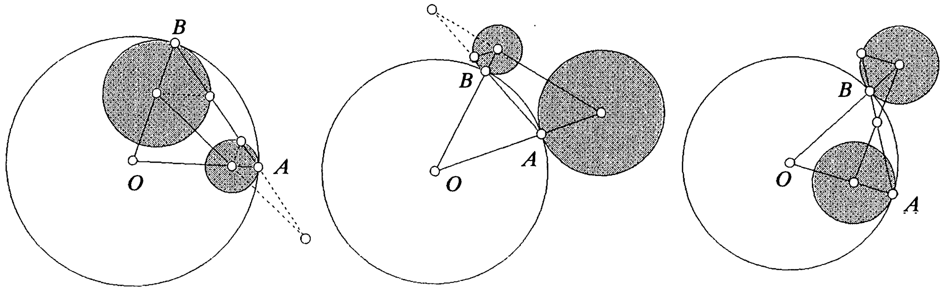


Figure 5

Here, we have made use of analogues of (2). Similarly, the external centers of similitude of the pairs of circles (K_C), (K_A), and (K_A), (K_B) are

$$Y = \frac{-\left(1 - \frac{r}{\rho_C}\right)C + \left(1 - \frac{r}{\rho_A}\right)A}{\left(\frac{1}{\rho_C} - \frac{1}{\rho_A}\right)r} \quad \text{and} \quad Z = \frac{-\left(1 - \frac{r}{\rho_A}\right)A + \left(1 - \frac{r}{\rho_B}\right)B}{\left(\frac{1}{\rho_A} - \frac{1}{\rho_B}\right)r}.$$

These three points X, Y, Z all lie on the line

$$\frac{x}{1 - \frac{r}{\rho_A}} + \frac{y}{1 - \frac{r}{\rho_B}} + \frac{z}{1 - \frac{r}{\rho_C}} = 0. \tag{5}$$

Indeed, the triangles ABC, $A'B'C'$, and $K_A K_B K_C$ are pairwise in perspective, with line (5) as common axis of perspective.

We close with a few remarks. Since the points X, Y, Z are the external centers of similitude of pairs of circles from (K_A), (K_B), (K_C), their collinearity also follows from the famous Desargues Three-Circle Theorem [**5**]. If we make use of (1), this axis of perspective has equation

$$\frac{x}{\sin^2\frac{\alpha}{2}} + \frac{y}{\sin^2\frac{\beta}{2}} + \frac{z}{\sin^2\frac{\gamma}{2}} = 0.$$

Finally, we note another interesting consequence of (1). The Gergonne point of a triangle is the point of intersection of the three cevians joining each vertex to the point of contact of the incircle with the opposite side. This is the point X_7 of [**4**], and has trilinear coordinates

$$\sec^2\frac{\alpha}{2} : \sec^2\frac{\beta}{2} : \sec^2\frac{\gamma}{2}.$$

As such, this is the unique point whose distances to the sides are proportional to the radii of the mixtilinear incircles in the respective angles.

REFERENCES

1. L. Bankoff, A Mixtilinear Adventure, *Crux Math.* **9** (1983) 2–7.
2. C. V. Durell and A. Robson, *Advanced Trigonometry*, Bell & Sons, 1935.
3. R. A. Johnson, *Advanced Euclidean Geometry*, Dover reprint, 1965.
4. C. Kimberling, Central Points and Central Lines in the Plane of a Triangle, *Math. Mag.* **67** (1994) 163–187.
5. J. McCleary, An Application of Desargues' Theorem, *Math. Mag.* **55** (1982) 233–235.

Florida Atlantic University, Boca Raton, FL 33431
yiu@fau.edu

The Area of the Medial Parallelogram of a Tetrahedron

David N. Yetter

The midpoints of any four edges of a Euclidean tetrahedron that form a cycle are coplanar, and are the vertices of a parallelogram. The purpose of this note is to derive a simple formula for the area of this *medial parallelogram* of a tetrahedron in terms of the lengths of the six edges. It would appear that this result is either new or long-forgotten.

Despite the very classical nature of the problem our formula solves, there is some serious contemporary interest arising from recently proposed simplicial models for quantum gravity, in which such a formula is needed to approach the problem of length operators; see **[1]**, **[2]**.

Consider a tetrahedron with edge-lengths as in Figure 1. Fix a pair of non-incident edges, say those of lengths e and f. It is then easy to see that the midpoints of the remaining four edges lie in a plane parallel to both of the chosen edges, and equidistant from the planes containing each chosen edge and parallel to both, and that they form the vertices of a parallelogram in this plane.

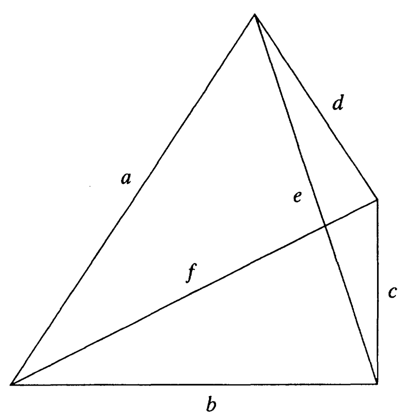


Figure 1. A generic tetrahedron

Definition 1. Given a pair of non-incident edges in a tetrahedron, the *medial parallelogram* determined by the pair is the parallelogram whose vertices are the midpoints of the remaining four edges.

Our main result is

Theorem 2. *The area of the medial parallelogram determined by the edges of lengths e and f in the tetrahedron of Figure 1 is*

$$\frac{1}{8}\sqrt{4e^2f^2 - (a^2 - b^2 + c^2 - d^2)^2}$$

Proof: The key is to consider the vertices of the tetrahedron as vectors $2\vec{p}$, $2\vec{q}$, $2\vec{r}$, and $2\vec{s}$ in $\mathbb{R}^3$. The factors of 2 in the vertices given as vectors are included to avoid fractions; see Figure 2.

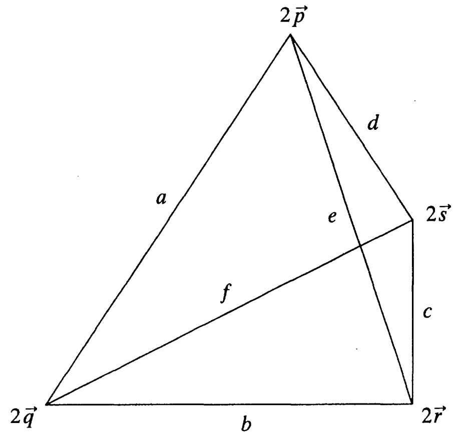


Figure 2. Tetrahedron with vertices as vectors

The vertices of the medial parallelogram are then given by the vectors $\vec{p}+\vec{q}$, $\vec{q}+\vec{r}$, $\vec{r}+\vec{s}$, and $\vec{s}+\vec{p}$. The lengths of the six edges are given in terms of the six vectors by

$$a = 2|\vec{p}-\vec{q}|, \quad b = 2|\vec{q}-\vec{r}|, \quad c = 2|\vec{r}-\vec{s}|,$$
$$d = 2|\vec{s}-\vec{p}|, \quad e = 2|\vec{r}-\vec{p}|, \quad f = 2|\vec{s}-\vec{q}|.$$

The medial tetrahedon is then spanned by the vectors $\vec{u}=\vec{r}-\vec{p}$ and $\vec{v}=\vec{s}-\vec{q}$; see Figure 3.

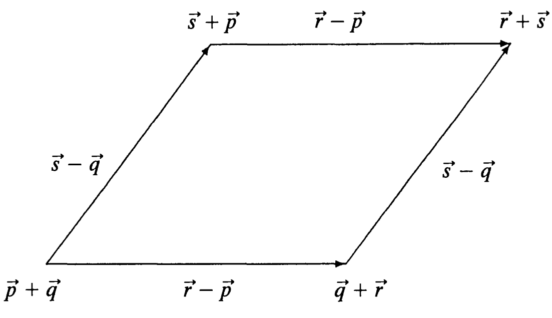


Figure 3. The medial tetrahedron in terms of vectors

The area of the medial tetrahedron is thus $|\vec{u}\times\vec{v}|$.

Now, recall that since $\sin^2\theta = 1-\cos^2\theta$, the vector (cross) and scalar (dot) products of an two vectors $\vec{x}$ and $\vec{y}$ in $\mathbb{R}^3$ are related by

$$|\vec{x}\times\vec{y}|^2 = |\vec{x}|^2|\vec{y}|^2 - (\vec{x}\cdot\vec{y})^2.$$

thus, in our case we have

$$
\begin{aligned}
\text{Area}^2 &= |\vec{u} \times \vec{v}|^2 \\
&= |\vec{r}-\vec{p}|^2 |\vec{s}-\vec{q}|^2 - \left[(\vec{r}-\vec{p})\cdot(\vec{s}-\vec{q})\right]^2 \\
&= \frac{1}{16}e^2 f^2 - \left[\vec{r}\cdot\vec{s} - \vec{r}\cdot\vec{q} - \vec{p}\cdot\vec{s} + \vec{p}\cdot\vec{q}\right] \\
&= \frac{1}{16}e^2 f^2 - \frac{1}{4}\left[-2\vec{r}\cdot\vec{s} - (-2)\vec{r}\cdot\vec{q} - (-2)\vec{p}\cdot\vec{s} + (-2)\vec{p}\cdot\vec{q}\right]^2 \\
&= \frac{1}{16}e^2 f^2 - \frac{1}{4}\Big[\left(|\vec{r}|^2 - 2\vec{r}\cdot\vec{s} + |\vec{s}|^2\right) - \left(|\vec{r}|^2 - 2\vec{r}\cdot\vec{q} + |\vec{q}|^2\right) \\
&\qquad -\left(|\vec{p}|^2 - 2\vec{p}\cdot\vec{s} + |\vec{s}|^2\right) + \left(|\vec{p}|^2 - 2\vec{p}\cdot\vec{q} + |\vec{q}|^2\right)\Big]^2 \\
&= \frac{1}{16}e^2 f^2 - \frac{1}{4}\left[|\vec{r}-\vec{s}|^2 - |\vec{r}-\vec{q}|^2 - |\vec{p}-\vec{s}|^2 + |\vec{p}-\vec{q}|^2\right]^2 \\
&= \frac{1}{16}e^2 f^2 - \frac{1}{4}\left[\left(\frac{c}{2}\right)^2 - \left(\frac{b}{2}\right)^2 - \left(\frac{d}{2}\right)^2 + \left(\frac{a}{2}\right)^2\right]^2 \\
&= \frac{1}{64}\left[4e^2 f^2 - (a^2 - b^2 + c^2 - d^2)^2\right]
\end{aligned}
$$

Thus, taking square roots, we have the desired result. ■

ACKNOWLEDGMENTS. I thank R. Chapman, H.M.S. Coxeter, and T. Sudbery for suggestions that improved this Note, and for giving me confidence that the result is either new or long lost. I also thank the National Science Foundation for financial support under grant # DMS-9504423.

REFERENCES

1. Barbieri, A., Quantum tetrahedra and simplicial spin networks, e-print gr-qc/9707010.
   Barbieri, A., Space of vertices of relativistic spin networks, e-print gr-qc/9709076.
2. Barrett, J.W. and Crane, L., Relativistic spin networks and quantum gravity, e-print gr-qc/9709028.

Kansas State University, Manhattan, KS 66506
dyetter@math.ksu.edu

UNSOLVED PROBLEMS

Edited by **Richard Nowakowski**

In this department the MONTHLY presents easily stated unsolved problems dealing with notions ordinarily encountered in undergraduate mathematics. Each problem should be accompanied by relevant references (if any are known to the author) and by a brief description of known partial or related results. Typescripts should be sent to Richard Nowakowski, Department of Mathematics & Statistics, Dalhousie University, Halifax NS, Canada B3H 3J5, rjn@mscs.dal.ca

Unsolved Problems, 1969–1999

References in brackets are to year and page numbers of this MONTHLY, while dates in parentheses refer to publications listed at the end; other items are labelled (tbp) if they are likely to be published formally, or as written communications (wrc) if publication plans are not now known. Dates and pages in brackets are also appended to items in the bibliography indicating where the problem originally appeared in the MONTHLY.

Sommers (1998) gives some convex solutions to the sofa problem [1976, 188] and the Erikssons (1998) treat rectangular food-trolleys going round corners of any angle between corridors of different widths. A reference not made earlier, and not in **G5** of Croft, Falconer, and Guy (1991) is Davenport (1986).

In spite of exhortations [1983, 36], people continue to attempt to solve the $3x + 1$ problem. If we iterate the function $T(n) = n/2$ (n even), $(3n + 1)/2$ (n odd), then we can define the *stopping time*, $s(n)$, as the least number k of iterations that give $T^k(n) < n$, and the *maximum excursion*, $t(n)$, as the maximum value of $T^k(n)$ for $k > 0$. Are $s(n)$ and $t(n)$ always finite? Tomás Oliveira e Silva (1999) has verified the $3x + 1$ conjecture for $n \leq 3 \cdot 2^{53} \approx 2.702 \cdot 10^{16}$. He lists all the record holders for $s(n)$ and $t(n)$ in this range, the largest being $s(1008932249296231) = 886$ and $t(10709980568908647) = 175294593968539094415936960141122$. There is evidence that $t(n) < n^2 f(n)$ where $f(n)$ is either constant or very slowly increasing. The highest value found of $t(n)/n^2$ is 7.527 for $n = 3716509988199$. For only 7 of the 76 record-holders is the value greater than one.

There has been a good deal written about polynomials, such as $x^3 - 33x^2 + 216x$, all of whose derivatives have integer roots [1989, 129]; Buchholz and MacDougall (tbp) give 34 references. For quartics and quintics the situation is fully understood, except that it has not been proved that there are no such polynomials with four or more distinct roots, nor quintics with three distinct roots, one of them triple.

Coxeter (1989) solved his 'challenging definite integral' [1988, 330] geometrically, while Peter Wagner (1996) includes an analytical solution in proving the more general result,

$$\frac{5}{2}\int_{\frac{1}{2}\cot^2\alpha}^{1} \frac{\arccos x}{(2x+1)\sqrt{x+1}}\left[\frac{\sqrt{2}\cos\alpha}{\sqrt{2x\sin^2\alpha-\cos 2\alpha}}+\frac{2}{\sqrt{x}}\right]dx$$

$$= \int_{\alpha_0}^{\alpha} \arccos\left(\frac{\cos 2\alpha(\sin^2 t+\cos 2\alpha)}{\sin^2 t-\cos^2 2\alpha}\right)dt$$

$$- \int_{\alpha_1}^{\alpha} \arccos\left(\frac{\cos 2t}{1-2\cos 2t}\right)dt + 4\pi\left(\alpha-\frac{\pi}{4}\right)_+$$

valid for $\alpha_1 \le \alpha \le \pi/2$, where $\alpha_0 = \arccos(\cot\alpha\sqrt{1-2\cos 2\alpha})$, $\alpha_1 = \operatorname{arccot}\sqrt{2}$, and $\left(\alpha - \frac{\pi}{4}\right)_+$ is Heaviside's function, namely $\alpha - \frac{\pi}{4}$ for $\alpha \ge \frac{\pi}{4}$ and 0 for $\alpha \le \frac{\pi}{4}$.

Shattuck and Cooper (tbp) have found divergent RATS sequences [1989, 425] in bases 50, 99, 148, 962, $18n+1$, $18n+10$, and $(2^t-1)^2+1$, where t is a prime or pseudoprime, base 2. Conway's conjecture, that in base 10, all RATS (Reverse, Add, Then Sort) sequences either cycle or are tributary to the sequence

$$1\,2\,3^m\,4^4\,5^2\,6^m\,7^4,\ 1\,2\,3^{m+1}\,4^4\,5^2\,6^{m+1}\,7^4, \ldots,$$

remains an open question.

Scott Hochwald corrected [1993, 947] a result of Tony Gardiner [1988, 927] and now has further results. Let $S(n,p) = \sum_{k=1}^{p-1}(k^n)$ and $H(n,p) = \sum_{k=1}^{p}(1/k^n)$. Let $A(n)$ be the set of primes, p, such that the numerator of $H(p^n, p-1)$ is divisible by p^{n+3} and let $B(n)$ be the set of primes, p, such that $S(p^{2n+1}-p^{2n}-p^n, p)$ is divisible by p^{n+3}. Gardiner showed that

$$\{\text{primes } p : \text{the numerator of } H(1, p-1) \text{ is divisible by } p^3\}$$
$$= \{\text{primes } p : \text{ the numerator of } H(2, p-1) \text{ is divisible by } p^2\}.$$

Hochwald′ has shown that these two sets are equal to $A(n)$ and $B(n)$ for $n = 1, 2, 3, 4, \ldots$.

Also, he has shown that if p is a prime larger than 3, and if m and n are positive integers chosen so that m is not divisible by $p-1$, then the numerator of $H(mp^n, p-1)$ is divisible by p^{n+1}; and the numerator of $H(mp^n, p-1)$ is divisible by p^{n+2} whenever m is odd and $m+1$ is not divisible by $p-1$.

Further coin-weighing [1995, 164] results where given by Wan and Du (1997). Suppose there are n coins of which d are light. Let $M_A(n; d)$ ($M_A(n, d)$) denote the maximum number of tests needed by an algorithm A to sort n coins where the number, d, of light coins is unknown (known). The algorithm A has a competitive ratio of c if there is a constant b such that for all $0 < d < n$, $M_A(n; d) \le c \cdot M_A(n, d) + b$. Wan et al. found an algorithm with competitive ratio $1/2 + \ln 3$. This improves on the earlier values, $(3/2)\ln 3$, $2\ln 3$, and $3\ln 3$, of c found by Wan, Yang, and Kelley (1997), Hu and Hwang (1994), and Hu, Chen, and Hwang (1995).

Andrew Bremner (tbp) has continued the search for a 3×3 magic square whose entries are distinct squares [1995, 925]. He approaches the problem from two different directions: to find a magic square with a maximum number of square

entries; to find a square with square entries and a maximum number of magic sums. He gives a parametric solution with 7 of the 8 magic sums equal, but can achieve a truly magic square only in fields of degrees $4, 8, 16, 20, 24, 27, 28, 32, 34, \ldots$. Lee Sallows (1997) has a relevant article.

In writing about the problem [1997, 359] of finding solutions to the equation $\phi(n) + \sigma(n) = kn$ proposed by Zhang, Lin, and Wang, where $\sigma(n)$ is the sum of divisors function and $\phi(n)$ is Euler's totient function, we omitted C. A. Nicol's (1966) paper in which he shows that if $k \geq 3$, then n is not squarefree, and if k is odd, then n is even or the square of an odd composite integer. He also shows that, for $k = 3$, if $q = 7 \cdot 2^{r-2} - 1$ is prime, then $n = 2^r \cdot 3q$ is a solution; this is so for $r =$ 3, 7, 11, 19, 23, 31, 47, 179, 18383, 22531, 24559, 26111, 34859, 41959, 67423, and 70211, but no one is likely to show that this gives an infinity of solutions. In a 98-04-26 email, James Ordway sent the solution $n = 2^6 \cdot 3 \cdot 113 \cdot 6343$ for the case $k = 3$.

Irving Kaplansky notes that not every integer is the sum of three cubes, as was asked in [1998, 953], but those that are not $\pm 4 \bmod 9$ may be.

In describing the 'greedy odd algorithm' for Egyption fractions [1998, 953], it should have been made clear that the $1/n$ that was to be subtracted from a given rational number should have the smallest **odd** n that left a non-negative remainder. The question is, does repetition of the process always lead to a zero remainder? A more spectacular example, $2/24631$, was found by Broadhurst; the numerators are

$$2, 3, 4, 5, 6, \ldots, 25, 2, 3, 4, 5, 2, 1.$$

Even more spectacular examples were found by Broadhurst:

fraction	with	numerator
$2/588391$	28	$2 \ldots 28\,1$
$4/538199$	28	$4 \ldots 25\,2 \ldots 5\,2\,1$
$3/46547$	29	$3 \ldots 25\,2 \ldots 5\,2\,1$
$2/24631$	30	$2 \ldots 25\,2 \ldots 5\,2\,1$
$6/104651$	30	$6 \ldots 32\,3\,2\,1$
$5/1444613$	37	$5\,2 \ldots 34\,43\,64\,1$

where the last denominator is a number of 384122451172 digits. David Eppstein (wrc) mentions that $7/111392341457976533660423$ also has a long expansion.

Kevin Brown has a method for constructing fractions with arbitrarily long odd greedy expansions at
`http://www.seanet.com/ ksbrown/kmath478.htm.`

Gary Mulkey (wrc) and Tom Hagedorn [tbp] have each proved the Hardin-Sloane conjecture [1998, 953] that if $n > 3$ is odd and not a multiple of 3, then $3/n$ can be expressed as the sum of the reciprocals of three distinct odd positive integers.

Marc Paulhus [1999, 162] should have referred to Beasley (1989), who devoted four pages to Beggar-My-Neighbour, giving a computer simulation and a probabilistic heuristic that there is at least a 90% of there being a loop in the game, but noting the common feature of many combinatorial problems, that, as the numbers increase, the size of the haystack increases exponentially relative to the size of the needle we're looking for. Our interest in the problem was restimulated by independent enquiries from Reg Allenby and John Mackay; further interest may be generated by the recent television production of *Great Expectations*.

We are indebted to numerous correspondents for help with this compilation.

John D. Beasley, *The Mathematics of Games*, Oxford University Press, 1989, esp. pp. 149–153. [1999, 162]

Andrew Bremner, On squares of squares, *Acta Arith.* (to appear). [1995, 925]

Ralph H. Buchholz and James A. MacDougall, Rational-derived polynomials and their extensions to quadratic fields, preprint. [1989, 129]

H. S. M. Coxeter, Trisecting an orthoscheme; symmetry 2: unifying human understanding, Part 1, *Comput. Math. Appl.* **17** (1989) 59–71; *MR* **90d**:51034. [1988, 330]

H. T. Croft, K. J. Falconer, and R. K. Guy, *Unsolved Problems in Geometry*, Springer-Verlag, New York, 1991, §**G5**. [1976, 188]

J. H. Davenport, A "piano movers" problem, *SIGSAM* **76** (1986) 15–17. [1976, 188]

Gerd Eriksson, Henrik Eriksson and Kimmo Eriksson, Moving a food trolley around a corner, *Theoret. Comput. Sci.* **191** (1998) 193–203. [1976, 188]

Thomas R. Hagedorn, A proof of the Hardin-Sloane conjecture, preprint. [1998, 953]

X.-D. Hu, P.-D. Chen, and F. K. Hwang, A new competitive algorithm for the counterfeit coin problem, *Inform. Process Lett.* **51** (1994) 213–218. [1995, 164]

X.-D. Hu and F. K. Hwang, A competitive algorithm for the counterfeit coin problem, in *D.-Z. Du and P. M. Pardalos, eds.*, Minimax and Applications, Kluwer Academic Publishers, 1995, pp. 241–250. [1995, 164]

C. A. Nicol, Some diophantine equations involving arithmetic functions, *J. Math. Anal. Appl.* **15** (1966) 154–161; *MR* **33** #4007. [1997, 358]

Lee Sallows, The lost theorem, *Math. Intelligencer* **19** (1997) 51–54. [1995, 925]

Tomás Oliveira e Silva, Maximum excursion and stopping time record-holders for the $3x + 1$ problem: computational results, *Math. Comput.* **68** (1999) 371–384. [1983, 36]

Steven Shattuck and Curtis Cooper, Divergent RATS sequences, *Proc. Ninth Internat. Conf. Fibonacci Numbers Appl.*, Luxembourg, 2000 (to appear). [1989, 425]

James A. Sommers, On some convex solutions to the sofa problem (98-05-25 preprint). [1976, 188]

Peter Wagner, Solution to a problem posed by H. S. M. Coxeter, *C. R. Math. Rep. Acad. Sci. Canada* **18** (1996) 273–277. [1988, 330]

Wan Peng-Jun, Yang Qi-Fan, and Dean Kelley, A $(3/2)\log 3$-competitive algorithm for the counterfeit coin problem, *Theor. Comput. Sci.* **181** (1997) 347–356. [1995, 164]

Wan Peng-Jun and Du Ding-Zhu, A $(\log_2 3 + 1/2)$-competitive algorithm for the counterfeit coin problem, *Discrete Math.* **163** (1997) 173–200. [1995, 164]

PROBLEMS AND SOLUTIONS

Edited by **Gerald A. Edgar, Daniel H. Ullman,** and **Douglas B. West**

with the collaboration of Paul T. Bateman, Mario Benedicty, Paul Bracken, Duane M. Broline, Ezra A. Brown, Richard T. Bumby, Glenn G. Chappell, Randall Dougherty, Roger B. Eggleton, Ira M. Gessel, Bart Goddard, Jerrold R. Griggs, Douglas A. Hensley, John R. Isbell, Robert Israel, Kiran S. Kedlaya, Murray S. Klamkin, Fred Kochman, Frederick W. Luttmann, Vania Mascioni, Frank B. Miles, Richard Pfiefer, Cecil C. Rousseau, Leonard Smiley, John Henry Steelman, Kenneth Stolarsky, Richard Stong, Charles Vanden Eynden, and William E. Watkins.

Proposed problems and solutions should be sent in duplicate to the MONTHLY *problems address on the inside front cover. Submitted problems should include solutions and relevant references. Submitted solutions should arrive at that address before May 31, 2000; Additional information, such as generalizations and references, is welcome. The problem number and the solver's name and address should appear on each solution. An acknowledgement will be sent only if a mailing label is provided. An asterisk (*) after the number of a problem or a part of a problem indicates that no solution is currently available.*

PROBLEMS

10767. *Proposed by Bruce Dearden and Jerry Metzger, University of North Dakota, Grand Forks, ND.* For integers $n \geq 2$ and $m > 1$, how many invertible m-by-m matrices are there modulo n?

10768. *Proposed by Sung Soo Kim, Hanyang University, Ansan, Kyunggi, Korea.*
(**a**) Show that there is a continuous function $f\colon \mathbb{R} \to \mathbb{R}$ such that $f + g$ is not increasing for any differentiable function g.
(**b**) Show that there is a differentiable function $f\colon \mathbb{R} \to \mathbb{R}$ such that $f + g$ is not increasing for any continuously differentiable function g.
(**c**) Show that, for any continuously differentiable function $f\colon \mathbb{R} \to \mathbb{R}$, there is a real analytic function g such that $f + g$ is increasing.

10769. *Proposed by Christian Blatter, Zürich, Switzerland.* Determine the minimum number of colors necessary to color the points of a sphere in such a way that points at spherical distance $\pi/2$ (i.e., points that subtend a right angle from the center of the sphere) get different colors.

10770. *Proposed by Călin Popescu, Louvain-la-Neuve, Belgium.* Suppose that m and n are integers with $1 < m < \phi(m) + n$, where $\phi(m)$ is the number of elements in $\{1, 2, \ldots, m\}$ that are relatively prime to m. Show that $\sum_{i=1}^{n}(-1)^i\binom{n}{i}i^m$ is divisible by m.

10771. *Proposed by Mowaffaq Hajja and Peter Walker, American University of Sharjah, Sharjah, U. A. E.* Evaluate $\int_0^1\int_0^1\int_0^1\left(1 + u^2 + v^2 + w^2\right)^{-2} du\, dv\, dw$.

10772. *Proposed by William C. Waterhouse, Pennsylvania State University, University Park, PA.* For any ordered field K, one can define the derivative of a function $f\colon K \to K$ as usual by $f'(x) = \lim_{y\to x}(f(y) - f(x))/(y - x)$. Suppose that every $f\colon K \to K$ with derivative identically zero is constant. Prove that K is isomorphic to the field of real numbers.

10773. *Proposed by Jean Anglesio, Garches, France.* Let $a_0, a_1, \ldots, a_k$ be positive integers. For $0 \le i \le k$, let p_i/q_i be the fraction in lowest terms with continued fraction expansion $[a_0, a_1, \ldots, a_i]$. Find the continued fraction expansions of

$$\sqrt{\frac{p_k\, p_{k-1}}{q_k\, q_{k-1}}}, \sqrt{\frac{p_k\, q_k}{p_{k-1}\, q_{k-1}}}, \sqrt{\frac{p_k^2 + p_{k-1}^2}{q_k^2 + q_{k-1}^2}}, \text{ and } \sqrt{\frac{p_k^2 + q_k^2}{p_{k-1}^2 + q_{k-1}^2}}$$

in terms of $a_0, a_1, \ldots, a_k$.

SOLUTIONS

Tracking the Incenters

10631 [1997, 975]. *Proposed by Greg Huber, University of Chicago, Chicago, IL.* Given a triangle T, let the *intriangle* of T be the triangle whose vertices are the points where the circle inscribed in T touches T. Given a triangle T_0, form a sequence of triangles $T_0, T_1, T_2, \ldots$ in which each T_{n+1} is the intriangle of T_n. Let d_n be the distance between the incenters of T_n and T_{n+1}. Find $\lim_{n\to\infty} d_{n+1}/d_n$ when T_0 is not equilateral.

Solution by the GCHQ Problems Group, Cheltenham, U. K. We show that $d_{n+1}/d_n \to 1/4$. Let A, B, C be the angles of a triangle, r its inradius, R its circumradius, and d the distance from its incenter to its circumcenter. Then

$$d^2 = R^2 - 2Rr \tag{1}$$

and

$$r = 4R\sin(A/2)\sin(B/2)\sin(C/2). \tag{2}$$

(H. S. M. Coxeter and S. L. Greitzer, *Geometry Revisited,* MAA, 1967). Now let A', B', C' be the angles of the intriangle of ABC (with A' on side BC, etc.). Then $A' = \pi/2 - A/2$, so

$$A' - \pi/3 = (-1/2)(A - \pi/3), \tag{3}$$

and similarly for B' and C'. From (3) we infer that triangle T_n approaches equilateral as $n \to \infty$. For the triangle T_n, with angles A_n, B_n, C_n, define $a_n = A_n - \pi/3$, $b_n = B_n - \pi/3$, $c_n = C_n - \pi/3$, and $S_n = a_n^2 + b_n^2 + c_n^2$. Then (3) implies that $S_{n+1}/S_n = 1/4$. Also, $a_n + b_n + c_n = 0$, so $(a_n + b_n + c_n)^2 = 0$, and therefore

$$S_n = -2(a_nb_n + b_nc_n + c_na_n). \tag{4}$$

Now define $U_n = 1 - 8\sin(A_n/2)\sin(B_n/2)\sin(C_n/2)$. Using (1) and (2) and observing that $R_{n+1} = r_n$, we obtain

$$\left(\frac{d_{n+1}}{d_n}\right)^2 = \frac{R_{n+1}^2}{R_n^2}\,\frac{U_{n+1}}{U_n} = 16\sin^2(A_n/2)\sin^2(B_n/2)\sin^2(C_n/2)\,\frac{U_{n+1}}{U_n}. \tag{5}$$

Note that

$$\begin{aligned} 2\sin(A_n/2) &= 2\sin(a_n/2 + \pi/6) = \sqrt{3}\,\sin(a_n/2) + \cos(a_n/2) \\ &= 1 + \frac{\sqrt{3}}{2}\,a_n - \frac{1}{8}\,a_n^2 + O(a_n^3). \end{aligned}$$

Therefore

$$\begin{aligned} U_n &= 1 - \left(1 + \frac{\sqrt{3}}{2}\,a_n - \frac{1}{8}\,a_n^2 + \cdots\right)\left(1 + \frac{\sqrt{3}}{2}\,b_n - \frac{1}{8}\,b_n^2 + \cdots\right)\left(1 + \frac{\sqrt{3}}{2}\,c_n - \frac{1}{8}\,c_n^2 + \cdots\right) \\ &= \frac{1}{8}\,S_n - \frac{3}{4}\big(a_nb_n + b_nc_n + c_na_n\big) + \text{ terms of degree 3 or higher} \\ &= \frac{1}{2}\,S_n + \text{ terms of degree 3 or higher,} \end{aligned}$$

by (4). Therefore $\lim_{n\to\infty} U_{n+1}/U_n = \lim_{n\to\infty} S_{n+1}/S_n = 1/4$. Putting $\lim_{n\to\infty} A_n = \lim_{n\to\infty} B_n = \lim_{n\to\infty} C_n = \pi/3$ into (5) yields $d_{n+1}^2/d_n^2 \to 1/16$, or $d_{n+1}/d_n \to 1/4$.

Solved also by J. Anglesio (France), G. L. Body (U. K.), R. J. Chapman (U. K.), J. E. Dawson (Australia), N. Lakshmanan, J. H. Lindsey II, P. G. Poonacha (India), V. Schindler (Germany), A. Tissier (France), and the proposer.

An Appearance of the Beta Function

10632 [1997, 975]. *Proposed by William F. Trench, Trinity University, San Antonio, TX.* For given nonnegative integers m and n, evaluate

$$\sum_{k=0}^{m} \frac{(-1)^k}{n+k+1}\binom{m}{k}(1-y)^{n+k+1} + \sum_{k=0}^{n} \frac{(-1)^k}{m+k+1}\binom{n}{k} y^{m+k+1}.$$

Solution by Ronald A. Kopas, Clarion University of Pennsylvania, Clarion, PA. The sum is $m!\,n!/(m+n+1)!$. To see this, note that

$$\begin{aligned}\int_0^y t^m(1-t)^n\,dt &= \int_0^y t^m \sum_{k=0}^{n}\binom{n}{k}(-1)^k t^k\,dt \\ &= \sum_{k=0}^{n}\binom{n}{k}(-1)^k \int_0^y t^{m+k}\,dt = \sum_{k=0}^{n}\binom{n}{k}(-1)^k \frac{y^{m+k+1}}{m+k+1}.\end{aligned}$$

Substituting $1-t$ for t and then computing in the same way yields

$$\int_y^1 t^m(1-t)^n\,dt = \int_0^{1-y} t^n(1-t)^m\,dt = \sum_{k=0}^{m}\binom{m}{k}(-1)^k \frac{(1-y)^{n+k+1}}{n+k+1}.$$

Hence the desired sum equals $\int_0^1 t^m(1-t)^n\,dt$, which repeated integration by parts reduces to $m!\,n!/(m+n+1)!$.

Editorial comment. Most solvers first differentiated the given expression to show that it was independent of y. They then evaluated the expression at $y = 0$ or $y = 1$ and got the final result either by induction or by reducing it to the beta integral that appears in the published solution.

Solved also by U. Abel (Germany), K. F. Andersen (Canada), P. J. Anderson (Canada), J. Anglesio (France), G. W. Arnold, G. Bach (Germany), D. Beckwith, J. C. Binz (Switzerland), G. L. Body (U. K.), P. Bracken (Canada), D. Callan, R. J. Chapman (U. K.), Q. H. Darwish (Oman), M. N. Deshpande (India), P. Devaraj & R. S. Deodhar (India), S. B. Ekhad, Z. Franco & M. Wood, R. García-Pelayo (Spain), C. Georghiou (Greece), T. Hermann, V. Hernández & J. Martín (Spain), D. Huang, G. Kesselman, M. S. Klamkin (Canada), R. A. Leslie, N. F. Lindquist, J. H. Lindsey II, S. McDonald & K. Adzievski, J. G. Merickel, C. A. Minh, D. A. Morales (Venezuela), R. G. Mosier, A. Nijenhuis, M. Omarjee (France), G. Peng, H. Qin, V. Schindler (Germany), H.-J. Seiffert (Germany), P. Simeonov, N. C. Singer, J. H. Steelman, R. F. Swarttouw (The Netherlands), A. Tissier (France), E. I. Verriest, M. Vowe (Switzerland), H. Widmer (Switzerland), M. Woltermann, Y. Yang, Q. Yao, Anchorage Math Solutions Group, BARC Problems Group (India), GCHQ Problems Group (U. K.), NSA Problems Group, WMC Problems Group, and the proposer.

Apéry's Constant

10635 [1998, 68]. *Proposed by Nicholas R. Farnum, California State University, Fullerton, CA.* Show that the value of $\zeta(s) = \sum_{k=1}^{\infty} k^{-s}$ at $s = 3$, also called Apéry's constant, can be expressed as $\zeta(3) = \sum_{n=1}^{\infty} r_n/n$, where $r_n = (\pi^2/6) - \sum_{k=1}^{n} k^{-2}$ is the nth remainder of the series expansion of $\zeta(2)$.

Solution by Alain Tissier, Montfermeil, France. We prove a generalization: For each positive integer k,

$$k!\,\zeta(k+2) = \sum_{n_1=1}^{\infty}\sum_{n_2=1}^{\infty}\cdots\sum_{n_k=1}^{\infty} \frac{1}{n_1 n_2 \ldots n_k} \sum_{p=1+n_1+n_2+\cdots+n_k}^{\infty} \frac{1}{p^2}. \qquad (*)$$

The case $k = 1$ gives the result of this problem, while the case $k = 2$ gives

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \frac{1}{nm} \left(\frac{\pi^2}{6} - \sum_{p=1}^{n+m} \frac{1}{p^2} \right) = \frac{\pi^4}{45}.$$

To prove $(*)$, observe that $\int_0^1 t^{p-1}(-\ln t)\,dt = p^{-2}$ for each positive integer p. This is the base case of a proof by induction that $\int_0^1 t^{p-1}(-\ln t)^m\,dt = m!/p^{m+1}$ for all positive integers p and m. Hence

$$\begin{aligned}
&\sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} \cdots \sum_{n_k=1}^{\infty} \frac{1}{n_1 n_2 \ldots n_k} \sum_{p=1+n_1+n_2+\cdots+n_k}^{\infty} \frac{1}{p^2} \\
&= \sum_{n_1=1}^{\infty} \sum_{n_2=1}^{\infty} \cdots \sum_{n_k=1}^{\infty} \frac{1}{n_1 n_2 \ldots n_k} \int_0^1 \frac{t^{n_1+n_2+\cdots+n_k}}{1-t} (-\ln t)\,dt \\
&= \int_0^1 \frac{(-\ln(1-t))^k}{1-t} (-\ln t)\,dt = \int_0^1 \frac{(-\ln s)^k}{s} (-\ln(1-s))\,ds \\
&= \frac{1}{k+1} \Big[(-\ln s)^{k+1} \ln(1-s) \Big]_0^1 + \frac{1}{k+1} \int_0^1 \frac{(-\ln s)^{k+1}}{1-s}\,ds \\
&= \frac{1}{k+1} \int_0^1 \frac{(-\ln s)^{k+1}}{1-s}\,ds = \frac{1}{k+1} \sum_{n=1}^{\infty} \int_0^1 s^{n-1} (-\ln s)^{k+1}\,ds = k! \sum_{n=1}^{\infty} \frac{1}{n^{k+2}}.
\end{aligned}$$

Editorial comment. A closely related identity appeared as Problem 4431 [1951, 195; 1952, 471] in this MONTHLY. A generalization appeared as Problem 1302 in *Math. Mag.* **62** (1989) 275: For each integer $n \geq 3$, $\zeta(n) = \sum_{i=1}^{n-2} \sum_{p,q=1}^{\infty} p^{-i}(p+q)^{i-n}$.

Readers pointed out a large number of references to this problem and various generalizations, going back to work of Euler in 1743. Among these references were: W. E. Briggs, S. Chowla, A. J. Kempner, and W. E. Mientka, On some infinite series, *Scripta Math.* **21** (1955) 28; J. M. Borwein and R. Girgensohn, Evaluating triple Euler sums, *Electronic J. Comb.* **3** (1996) R23; and B. C. Berndt, *Ramanujan's Notebooks*, Part I (1985) Springer-Verlag, p. 252.

Solved also by P. J. Andersen (Canada), J. Anglesio (France), D. & J. Borwein (Canada), P. Bracken (Canada), D. M. Bradley (Canada), D. Callan, R. J. Chapman (U. K.), H. Chen, C. Georghiou (Greece), W. Janous (Austria), P. Khalili, V. Lucic (Canada), J. G. Merickel, S. Northshield, L. Quet, V. Schindler (Germany), H.-J. Seifert (Germany), M. Sharma (India), P. Simeonov, I. Sofair, A. Stenger, T. V. Trif (France), D. B. Tyler, J. J. van Lint (Netherlands), M. Vowe (Switzerland), GCHQ Problems Group (U. K.), and the proposer.

Essentially Discontinuous Functions

10668 [1998, 465]. *Proposed by Abram Kagan, University of Maryland, College Park, MD, and Larry Shepp, Rutgers University, New Brunswick, NJ.* Let H be an infinite-dimensional closed subspace of $L^2[0, 1]$. Prove that H contains a function f that is essentially discontinuous, meaning that there is no continuous function g on $[0, 1]$ equal to f almost everywhere. Does the conclusion remain true if g is required to be continuous only on $(0, 1]$?

Solution by Kenneth Schilling, University of Michigan, Flint, MI. Suppose, for purposes of contradiction, that H contains no essentially discontinuous function. Then, since every continuous function on $[0, 1]$ is bounded, every element of H is essentially bounded. Let $\|\cdot\|_2$ and $\|\cdot\|_\infty$ denote the L^2 and essential supremum norms, respectively. Since $\|f\|_2 \leq \|f\|_\infty$, the identity map from the Banach space $(H, \|\cdot\|_2)$ to the Banach space $(H, \|\cdot\|_\infty)$ is continuous. By the Open Mapping Theorem, the inverse function is also continuous, so there exists $K > 0$ such that $\|f\|_\infty \leq K \cdot \|f\|_2$ for all $f \in H$.

Now let $f_1, \ldots, f_n$ be continuous functions that are orthonormal in H. For all real numbers $a_1, \ldots, a_n$ and all $x \in [0, 1]$, we have

$$\sum_{i=1}^{n} a_i f_i(x) \le K \cdot \left\| \sum_{i=1}^{n} \alpha_i f_i \right\|_2 = K \sqrt{\sum_{i=1}^{n} \alpha_i^2}.$$

Fix $x \in [0, 1]$, and let $\alpha_i = f_i(x)$. Then $\sum_{i=1}^{n} (f_i(x))^2 \le K \cdot \sqrt{\sum_{i=1}^{n} (f_i(x))^2}$, so $\sum_{i=1}^{n} (f_i(x))^2 \le K^2$. Integrating both sides from 0 to 1 gives $n \le K^2$. Thus every orthonormal set of continuous functions in H has at most K^2 elements. This contradicts the assumption that H is infinite-dimensional.

The conclusion does not follow with $(0, 1]$ in place of $[0, 1]$. For $n = 1, 2, \ldots$, let $f_n \colon [0, 1] \to \mathbf{R}$ be a continuous function with $\|f_n\|_2 = 1$ and support in $(1/(n+1), 1/n)$. Then $\{f_n\}$ is an orthonormal set, so the map $\Phi\colon l^2 \to L^2[0, 1]$ given by $\Phi(\alpha) = \sum_{n=1}^{\infty} \alpha_n f_n$ is a linear isometry. In addition, each $\Phi(\alpha)$ is continuous on $(0, 1]$, since for all $x \in (0, 1]$ there exists an open interval I about x such that $f_n \neq 0$ on I for at most one n. Thus the range of Φ is a closed, infinite-dimensional subspace of $L^2(0, 1]$ whose elements are continuous functions.

The first part of this problem is contained in problems 28 and 55 in Chapter 10 of H. L. Royden, *Real Analysis*, Third Edition, Macmillan, 1988. The solution here follows Royden's generous hints.

Solved also by P. J. Fitzsimmons, P. M. Jarvis, J. H. Lindsey II, A. Sasane (The Netherlands), and the proposers.

Two Recurrence Relations, One Easy, One Hard

10670 [1998, 559]. *Proposed by Salomon Benchimol and Elliott Cohen, Paris, France.*
(a) For which values of $u_0 > 0$ and $u_1 > 0$ does the sequence defined by $u_{n+2} = 1 + u_{n+1}/u_n$ for $n \ge 0$ converge?
(b) For which values of $u_0 > 0$ and $u_1 > 0$ does the sequence defined by $u_{n+2} = 1 + u_n/u_{n+1}$ for $n \ge 0$ converge?

Solution of part **(a)** *by Con Amore Problems Group, Copenhagen, Denmark.* This sequence converges to 2 for every choice of $u_0, u_1 > 0$. Clearly $u_n > 0$ for all n, so $u_n = 1 + u_{n-1}/u_{n-2} > 1$ for $n \ge 2$. If $n \ge 5$, then $u_n = 1 + u_{n-1}/u_{n-2} = 1 + 1/u_{n-3} + 1/u_{n-2} < 3$. This proves the $k = 0$ case of the following claim: For any $k \ge 0$,

$$u_n > \frac{2^{2k+2} - 1}{2^{2k+1} + 1} \text{ for } n \ge 6k + 2, \qquad \text{and} \qquad u_n < \frac{2^{2k+3} + 1}{2^{2k+2} - 1} \text{ for } n \ge 6k + 5.$$

This proves convergence, since both of these bounds converge to 2 as $k \to \infty$. We prove the claim by induction. Choose $k \ge 1$ and assume that the claim holds for smaller values of k. For $n \ge 6(k-1) + 5 = 6k - 1$, we have

$$u_n < \frac{2^{2(k-1)+3} + 1}{2^{2(k-1)+2} - 1} = \frac{2^{2k+1} + 1}{2^{2k} - 1}.$$

Therefore, for $n \ge 6k + 2$, we have

$$u_n = 1 + \frac{1}{u_{n-2}} + \frac{1}{u_{n-3}} > 1 + 2\,\frac{2^{2k} - 1}{2^{2k+1} + 1} = \frac{2^{2k+2} - 1}{2^{2k+1} + 1},$$

as required. For $n \ge 6k + 5$, we then have

$$u_n = 1 + \frac{1}{u_{n-2}} + \frac{1}{u_{n-3}} < 1 + 2\,\frac{2^{2k+1} + 1}{2^{2k+2} - 1} = \frac{2^{2k+3} + 1}{2^{2k+2} - 1},$$

as required.

Editorial comment. No correct solutions of **(b)** were received. It appears that the set of pairs (x, y) such that the sequence defined by $u_0 = x, u_1 = y, u_{n+2} = 1 + u_n/u_{n+1}$ converges is a curve through (2, 2) of the form

$$y = 2 + \frac{1}{2}(x-2) - \frac{1}{20}(x-2)^2 + \frac{7}{600}(x-2)^3 - \frac{71}{20400}(x-2)^4 + \cdots.$$

Part **(a)** solved also by S. S. Kim and the proposer.

The Number of Zeros of a Maclaurin Polynomial

10671 [1998, 559]. *Proposed by F. Rothe, University of North Carolina, Charlotte, NC.* Let

$$P_n(x) = \sum_{k=0}^{n} (-1)^k \frac{x^{2k+1}}{(2k+1)!}$$

be the Maclaurin polynomial of order $2n + 1$ of the sine function. Let c_n be the number of real zeros of P_n. Determine $\lim_{n\to\infty} c_n/(2n + 1)$.

Composite solution by Sung Soo Kim, Hanyang University, Ansan, Kyunggi, Korea, and the editors. The integral form of Taylor's theorem tells us that

$$P_n(x) = \sin x + \frac{(-1)^n}{(2n+1)!} e_{2n+1}(x), \qquad \text{where} \qquad e_k(x) = \int_0^x (x-t)^k \sin t \, dt.$$

Now $e_1(x) = x - \sin x$ and is positive for all $x > 0$, and $e_k'(x) = k e_{k-1}(x)$ for $k > 1$. Thus for $k \geq 3$, $e_k(x)$ is positive, increasing, and convex (concave up) on $(0, \infty)$.

Let $f_n(x) = e_{2n+1}(x)/(2n + 1)!$. We now consider the intervals $[a, b]$ on which $\sin x$ is monotone. Suppose first that n is even and $f_n(b) < 1$. If $a = (2m - 1/2)\pi$ and $b = (2m + 1/2)\pi$, then $P_n(x)$ is negative at a and positive at b and strictly increasing on $[a, b]$ so there is exactly one zero of P_n in $[a, b]$. If instead $a = (2m + 1/2)\pi$ and $b = (2m + 3/2)\pi$, then $P_n(x)$ is positive at a and negative at b. If $c = (2m + 1)\pi$, then P_n is positive on $[a, c]$. Thus P_n has at least one real zero in $[c, b]$. If there were more than one zero in $[c, b]$, there would have to be some $z \in [c, b]$ with $P_n''(z) < 0$: a convex function cannot be zero at more than one point on an interval if it is positive at one end and negative at the other. But $\sin(x)'' > 0$ on $[c, b]$, so also $P_n'' > 0$ on $[c, b]$, which is a contradiction. The case where n is odd is similar. The final case to be considered is when $f_n(a) < 1 < f_n(b)$. Here there can be two zeros in the interval, but again considerations of convexity forbid more.

This shows that the number of real zeros of P_n differs by at most a constant from the number of intervals $(k - \pi/2, k + \pi/2)$ in which $f_n < 1$. That number is given to within a bounded error by $2B(n)/\pi$, where $B(n)$ is the unique positive solution to $f_n(x) = 1$. But

$$e_k(x) = \int_0^x (x-t)^k \sin t \, dt < \int_0^\pi (x-t)^k \sin t \, dt \lesssim \pi x^k,$$

while

$$e_k(x) > \int_0^{2\pi} (x-t)^k \sin t \, dt = \sum_{j=0}^{k} \binom{k}{j} (x-\pi)^{k-j} \int_{-\pi}^{\pi} u^j \sin u \, du > 2\pi k (x-\pi)^{k-1}.$$

Thus $B(n)$ lies between the solutions to $x^{2n+1} = (2n+1)!/\pi$ and $(x-\pi)^{2n} = (2n)!/(2\pi)$. Both are asymptotically $2n/e$ by Stirling's formula, so $B(n) \approx 2n/e$. Thus, the number c_n of real zeros of $P_n(x)$ is asymptotic to $4n/e\pi$, so that $c_n/(2n + 1) \approx 2/e\pi$.

Editorial comment. David Bradley pointed out that the result is known and may be found (with details for the cosine function) in G. Szegő, Über eine Eigenschaft der Exponentialreihe, in *Gábor Szegő: Collected Papers 1915–1927*, Birkhauser, 1982, p. 659).

Solved also by J. H. Lindsey II, GCHQ Problems Group, and the proposer.

An AM–GM Variation

10672 [1998, 559]. *Proposed by V. Anil Kumar, Kerala Agricultural University, Tavanur, Kerala, India.* Let $p_1, p_2, \ldots, p_m$ be positive real numbers summing to 1, and assume that $a_{i,j} > 0$ for $1 \le i \le m$ and $1 \le j \le n$. Prove that

$$\sqrt[n]{\prod_{j=1}^{n}\left(\sum_{l=1}^{m} p_l a_{l,j} \prod_{i\neq l}\left(\sum_{k=1}^{n} a_{i,k}\right)\right)} \le \frac{1}{n}\prod_{i=1}^{m}\left(\sum_{k=1}^{n} a_{i,k}\right).$$

Solution by John H. Lindsey II, Fort Meyers, FL. With $x_j = \sum_{l=1}^{m} p_l a_{l,j} \prod_{i\neq l}(\sum_{k=1}^{n} a_{i,k})$, the left-hand side is the geometric mean of $x_1, \ldots, x_m$ and hence is less than or equal to the arithmetic mean of $x_1, \ldots, x_m$, which is

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}\left(\sum_{l=1}^{m} p_l a_{l,j} \prod_{i\neq l}\left(\sum_{k=1}^{n} a_{i,k}\right)\right) &= \frac{1}{n}\sum_{l=1}^{m} p_l \left(\sum_{j=1}^{n} a_{l,j} \prod_{i\neq l}\left(\sum_{k=1}^{n} a_{i,k}\right)\right)\\ &= \frac{1}{n}\sum_{l=1}^{m} p_l \prod_{i=1}^{m}\left(\sum_{k=1}^{n} a_{i,k}\right) = \frac{1}{n}\prod_{i=1}^{m}\left(\sum_{k=1}^{n} a_{i,k}\right).\end{aligned}$$

Solved also by S. Amighibech (France), R. J. Chapman (U. K.), Q. H. Darwish (Oman), W. Janous (Austria), B. Kalantari, S. S. Kim (Korea), M. S. Klamkin (Canada), R. Martin (U. K.), A. Nijenhuis, C. R. Pranesachar (India), H.-J. Seiffert (Germany), S. M. Soltuz (Romania), S.-E. Takahasi (Japan), T. V. Trif (Romania), GCHQ Problems Group (U. K.), and the proposer.

Functions with a Polynomial Addition Formula

10675 [1998, 560]. *Proposed by Harry Tamvakis, University of Pennsylvania, Philadelphia, PA.* Find every continuous function $f: \mathbb{R} \to \mathbb{R}$ such that some polynomial $P(x, y) \in \mathbb{R}[x, y]$ satisfies $f(x + y) = P\big(f(x), f(y)\big)$ for every $x, y \in \mathbb{R}$.

Solution by GCHQ Problems Group, Cheltenham, U. K. The function f can take one of two forms:

(i) $f(x) = ax - c$ using $P(u, v) = u + v + c$, including the special case of constant f when $a = 0$; and

(ii) $f(x) = (d^x - a)/b$ using $P(u, v) = a(u + v) + buv + (a^2 - a)/b$.

When $y = 0$, we get $f(x) = P(f(x), f(0)) = Q(f(x))$ for some polynomial Q. If the degree of Q is more than 1, then the value of f is restricted to the roots of the polynomial $Q(f) - f = 0$. Since f is continuous, it must be constant.

Assume now that the degree of Q is 1 and f is not constant. Since $f(x+y) = f(y+x)$, $P(u, v)$ is symmetric in u and v and must be of the form $a(u+v)+buv+c$. Setting $y = 0$ yields

$$f(x) = P(f(x), f(0)) = a(f(x) + f(0)) + bf(0)f(x) + c,$$

so $f(x)(1-a-bf(0)) = af(0)+c$. Since f is not constant, $1-a-bf(0) = 0 = af(0)+c$.

If $b = 0$, then $a = 1$ and $P(u, v) = u + v + c$. Hence $f(x + y) = f(x) + f(y) + c$, and so $f(0) = 2f(0) + c$ and $c = -f(0)$. Setting $g(x) = f(x) - f(0)$ yields $g(x + y) = g(x) + g(y)$ so that $g(x) = ax$ and $f(x) = ax - c$.

If $b \neq 0$, then $f(0) = (1 - a)/b = -c/a$, so $c = (a^2 - a)/b$. Hence $f(x + y) = a(f(x) + f(y)) + bf(x)f(y) + (a^2 - a)/b$, which yields

$$bf(x + y) + a = ab(f(x) + f(y)) + b^2 f(x)f(y) + a^2 = (bf(x) + a)(bf(y) + a).$$

Setting $g(x) = bf(x) + a$, we get $g(x + y) = g(x)g(y)$, and hence $g(x) = d^x$ for some $d > 0$. Thus $f(x) = (d^x - a)/b$.

Solved also by J. H. Lindsey II, A. Nijenhuis, and the proposer.

REVIVALS

An Unsettled Inequality

10337 [1993, 798; 1995, 659]. *Proposed by Horst Alzer, Waldbröl, Germany.* Let $n \geq 1$ be an integer. Let $x_1, \ldots, x_n$ be real numbers with $x_i \in (0, 1/2]$. Consider the statement

$$\prod_{i=1}^{n} \frac{x_i}{1 - x_i} \leq \frac{\sum_{i=1}^{n} x_i^n}{\sum_{i=1}^{n} (1 - x_i)^n}. \tag{$\mathbf{F}_n$}$$

(**a**) Prove $\mathbf{F}_n$ for $n \leq 3$.
(**b**) Show that $\mathbf{F}_n$ is false for $n \geq 6$.
(**c**)* What about $\mathbf{F}_4$ and $\mathbf{F}_5$?

Solution of part (**c**)* *by M. J. Pelling, London, England.* We show that $\mathbf{F}_4$ is true, with equality if and only if $x_1 = x_2 = x_3 = x_4$.

Write w, x, y, z for x_1, x_2, x_3, x_4, and write $\bar{w}, \bar{x}, \bar{y}, \bar{z}$ for $1 - w, 1 - x, 1 - y, 1 - z$, respectively. Then $\mathbf{F}_4$ may be written in the equivalent form

$$\frac{w^4 + x^4 + y^4 + z^4}{wxyz} \geq \frac{\bar{w}^4 + \bar{x}^4 + \bar{y}^4 + \bar{z}^4}{\bar{w}\bar{x}\bar{y}\bar{z}}. \tag{1}$$

Without loss of generality, suppose that $w \geq x \geq y \geq z$. Subtracting 4 from both sides of (1) and rearranging terms leads to

$$\frac{(w^2 - x^2)^2}{wxyz} + \frac{(y^2 - z^2)^2}{wxyz} + \frac{2(wx - yz)^2}{wxyz} \geq \frac{(\bar{w}^2 - \bar{x}^2)^2}{\bar{w}\bar{x}\bar{y}\bar{z}} + \frac{(\bar{y}^2 - \bar{z}^2)^2}{\bar{w}\bar{x}\bar{y}\bar{z}} + \frac{2(\bar{w}\bar{x} - \bar{y}\bar{z})^2}{\bar{w}\bar{x}\bar{y}\bar{z}}. \tag{2}$$

By repeated use of the elementary inequality

$$p + \frac{1}{p} \geq q + \frac{1}{q} \quad \text{whenever } p \geq q \geq 1, \tag{3}$$

we show that each term on the left of (2) is greater than or equal to the corresponding term on the right.

Since $w + x \leq 1$, we have $w - x \geq w^2 - x^2$ or $w\bar{w} \geq x\bar{x}$. With $p = w/x$ and $q = \bar{x}/\bar{w}$, we have $p \geq q \geq 1$, so

$$\frac{(w + x)^2}{wx} \geq \frac{(\bar{w} + \bar{x})^2}{\bar{w}\bar{x}} \tag{4}$$

by (3). Since $yz \leq \bar{y}\bar{z}$ and $(w - x)^2 = (\bar{w} - \bar{x})^2$, (4) implies

$$\frac{(w^2 - x^2)^2}{wxyz} = \frac{(w + x)^2}{wx}\frac{(w - x)^2}{yz} \geq \frac{(\bar{w} + \bar{x})^2}{\bar{w}\bar{x}}\frac{(\bar{w} - \bar{x})^2}{\bar{y}\bar{z}} = \frac{(\bar{w}^2 - \bar{x}^2)^2}{\bar{w}\bar{x}\bar{y}\bar{z}}. \tag{5}$$

The same reasoning proves

$$\frac{(y^2 - z^2)^2}{wxyz} \geq \frac{(\bar{y}^2 - \bar{z}^2)^2}{\bar{w}\bar{x}\bar{y}\bar{z}}. \tag{6}$$

Now let $p = wx/(yz)$ and $q = \bar{y}\bar{z}/(\bar{w}\bar{x})$. Again $p \geq q \geq 1$, so (3) implies

$$\frac{2(wx - yz)^2}{wxyz} \geq \frac{2(\bar{w}\bar{x} - \bar{y}\bar{z})^2}{\bar{w}\bar{x}\bar{y}\bar{z}}. \tag{7}$$

Adding (5), (6), and (7) yields (2).

Since equality holds in (3) only when $p = q$, we have equality in $\mathbf{F}_4$ only if $w/x = \bar{x}/\bar{w}$, $y/z = \bar{z}/\bar{y}$, and $wx/(yz) = \bar{y}\bar{z}/(\bar{w}\bar{x})$, which forces $w = x = y = z$.

Editorial comment. Pelling also contributed a lengthy proof of $\mathbf{F}_5$ and showed that equality holds in $\mathbf{F}_5$ only when $x_1 = x_2 = x_3 = x_4 = x_5$.

REVIEWS

Edited by **Harold P. Boas**
Mathematics Department, Texas A & M University, College Station, TX 77843-3368

Wavelets: A Primer. By Christian Blatter. A K Peters, 1998, x + 202 pp., $32.
Wavelets in a Box. By Charles K. Chui, Andrew K. Chan, and C. Steve Liu. Academic Press, 1998, book and CD-ROM software, $79.95.
A Primer on Wavelets for Scientists and Engineers. By James S. Walker. CRC Press, 1999, 155 pp., $39.95.
Wavelet Analysis: The Scalable Structure of Information. By Howard L. Resnikoff and Raymond O. Wells, Jr. Springer, 1998, xvi + 435 pp., $59.95.

Reviewed by **Edward Aboufadel, Matthew Boelkins, and Steven Schlicker**

In the decade since the publication of Ingrid Daubechies' seminal book **[6]**, wavelets have captured the imagination of an increasing number of people, and not just mathematicians. There is a growing enthusiasm about the subject on the part of students and many groups of professionals. Disciplines such as radiology, geology, computer science, music, and engineering provide a wide range of applications for wavelets, including signal and image processing, denoising of data, and compression and retrieval of data. Mathematicians continue to explore the area enthusiastically, as evidenced by well-attended sessions at national meetings, with papers on topics such as wavelets and dynamical systems and the path-connectivity of a wavelet space. The study of wavelets offers an intriguing mix of linear algebra, functional analysis, and applications, and it is time to let undergraduates in on the fun.

The abstract nature of much modern mathematics often leaves students and non-mathematicians asking the question, "Why does anyone care about this stuff?" Particularly in the undergraduate curriculum, the study of groups and rings, vector spaces, inner product spaces, topological spaces, and other abstract structures makes many students question the relevance of what they are learning. We mathematicians are engaged by the beauty and elegance of our subject, which often causes us to overlook the desire of most students to connect what they are studying directly to the world in which they live.

This need for context implies that part of our job as mathematics teachers is to demonstrate the power of mathematics through applications. By "applications", we do not mean ladders sliding down walls or the production of widgets, but rather *actual* uses of mathematics. True applications often require more mathematical background or sophistication than most students have, or more non-mathematical prerequisites than can be covered in class. Nonetheless, some applications that can be explained and demonstrated rely only on basic principles that are accessible to students. Wavelets are an excellent example of an application of high-level mathematics that can be presented to undergraduates.

Not long ago, as new wavelet enthusiasts, we became interested in how to present applications of wavelets to first- and second-semester linear algebra students. In reviewing the literature, we were disappointed to find few examples of papers or books written at a level accessible to typical undergraduates (or to non-mathematicians, for that matter). While nearly all the literature was in the style of Daubechies' graduate-level text, an article by Strang **[8]** showed us that the basic ideas can be reduced to linear combinations and matrix multiplication by focusing on the Haar wavelets. This idea, together with articles appearing in the popular press (**[5]** and **[7]**) and information on the FBI's interest in compressing fingerprint images (**[2]**, **[3]**, and **[4]**), served as our starting point for introducing wavelets to undergraduates.

This approach makes wavelets an appropriate topic for inclusion in a first-semester linear algebra course. By studying the FBI's fingerprint problem, our sophomore mathematics majors—many of whom intend to be high school teachers —have become excited about a new area of mathematics and have learned how to process simple two-dimensional images using the Haar wavelets. Adding a few additional topics (orthogonality and inner product spaces, which occur in a second course on linear algebra) makes wavelets appear in an even more natural and broader setting. These experiences convinced us that wavelets are significantly more accessible than most current books suggest.

As we refined our linear algebra projects on wavelets, we set up a web site **[1]** devoted to the topic of wavelets in the undergraduate curriculum. Since establishing the site in 1998, we have received many requests for more information about wavelets from students (both in mathematics and in other disciplines) and from non-mathematicians. The vast majority of people who contact us express frustration in trying to read the currently available literature on wavelets, most of which seems to be written for specialists.

Several recently published books purport, through title, advertising, or jacket notes, to be at an introductory level. Unfortunately, the word "introductory" is not well defined. In this review, we consider four such books and address their suitability for undergraduate students or non-mathematicians.

Wavelets in a Box is a great idea. In the package are a softcover copy of the book *An Introduction to Wavelets* by Charles K. Chui together with supporting software, *Wavelet Toolware*: *Software for Wavelet Training*, by C. Steve Liu and Andrew K. Chan. The intended audience for the book can be inferred from the first sentence of Section 1.1, which reads "Let $L^2(0,2\pi)$ denote the collection of all measurable functions f defined on the interval $(0,2\pi)$ with $\int_0^{2\pi}\|f(x)\|^2\,dx < \infty$." Most (probably all) undergraduates would hesitate to read any further. To comfort those daunted by this opening statement, the author offers these soothing words: "For the reader who is not familiar with the basic Lebesgue theory, the sacrifice is very minimal by assuming that f is a piecewise continuous function." By piecewise continuous the author means "...the existence of points $\{x_j\}$ in $\mathbf{R}$ with no finite accumulation points, such that $x_i < x_j$ for all j and that f is continuous on each of the open intervals (x_j, x_{j+1}) as well as the unbounded intervals $(-\infty, \min x_j)$ and $(\max x_j, \infty)$, if $\min x_j$ or $\max x_j$ exist." This is not easy reading for an undergraduate, or even an engineer, a geologist, or a radiologist.

Nevertheless, this is a nice book if one has the appropriate background. The author states that "the only prerequisite is a basic knowledge of function theory and real analysis." More realistically, what is needed is a strong background in real analysis, a little measure theory, at least enough complex analysis to understand complex exponential functions, and significant mathematical sophistication. The

lack of problems and examples also indicates that the book is at a higher level than an introduction. It includes a good primer on Fourier analysis, among other things, but this is a book written by a mathematician for mathematicians.

The companion software is a Windows-based package that can be used to process signals and images with wavelets. The accompanying guide states that *Wavelet Toolware* "is designed for the reader to gain some hands-on practice in the subject of wavelets." To a certain extent, this is correct. A student can use this package to create graphs of various famous scaling functions and mother wavelets through a built-in iterative process. One-dimensional signals can be processed via one-dimensional wavelet transformations with a choice of over a dozen different wavelet families. Students can also use the Continuous Wavelet Transform and the Short Time Fourier Transform tools. Since all of these tools are pre-coded, using the program requires almost no understanding of the mathematics of wavelets. This limits the educational value of the software.

We have found that image processing is an excellent way to motivate students, so we are glad to see that *Toolware* also contains a two-dimensional wavelet transform for the processing of grayscale images. The "2D DWT" tool reads files in binary PGM format (portable graymap, a creation of Jef Poskanzer) and creates image boxes, which are visual representations of wavelet coefficients, as shown in Figure 1. The choice of binary PGM is unfortunate since it is an uncommon format, used mostly on X Windows workstations. Consequently, students cannot easily create their own images to process. (This would have been possible if the

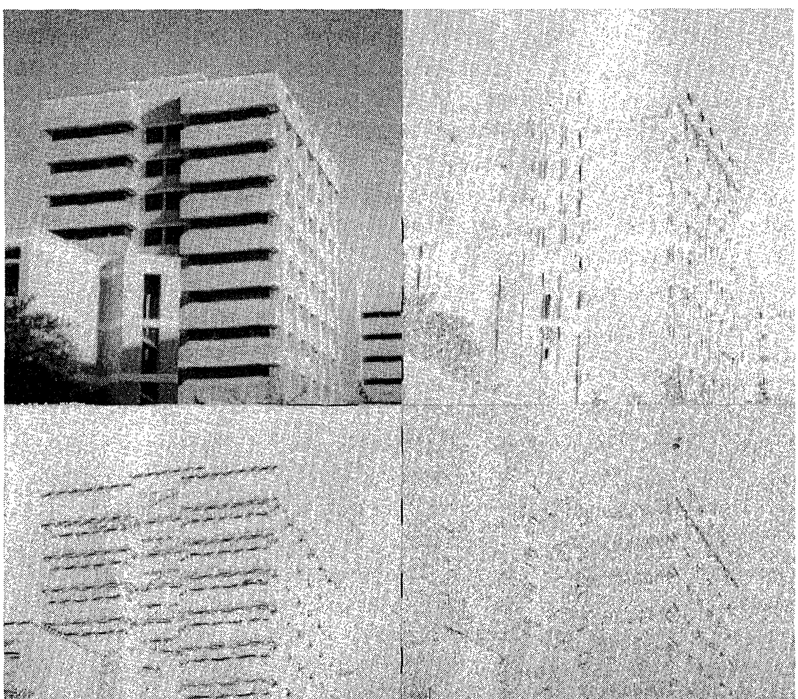


Figure 1. Image box from *Toolware*

raw PGM format had been used, since those files can be created with a text editor.) Instead, users must manipulate images from the campus of Texas A & M University, which are included in the package. This is great if you are an Aggie, but part of the fun of wavelets lies in working with one's own images. *Toolware* has a simple and straightforward user interface and would probably find its most useful role in providing in-class demonstrations.

Wavelets: A Primer is a promising title. According to the preface, the course from which the book arose was targeted towards "students of mathematics ... having the usual basic knowledge of analysis, carrying around a knapsack full of convergence theorems, but without any practical experience, say, in Fourier analysis." Actually, this slim volume belongs to the genre of wavelet books that require the reader to have a solid foundation in Fourier analysis to get past the first chapter. The author, Christian Blatter, does present some Fourier analysis in the second chapter, but the style is not very helpful to the novice.

The author remarks that this book is for students "in their senior year or first graduate year," and the latter category seems more accurate. The opening pages feature an overview of the problem of approximating functions, which is an improvement over books that begin, "A wavelet is... ." The writing is straightforward, and Blatter makes good use of figures. This text could serve as a resource for someone trying to read a book such as Chui's. Although *Wavelets: A Primer* is closer to the introductory level than other books of this type, the overall style and complete absence of problems still make this book too advanced for most undergraduate students.

Wavelet Analysis: The Scalable Structure of Information features a remarkable preface that describes the authors' history with a "mathematical engineering company" called Aware, Inc. As principal members of that company, the authors developed and implemented several ideas involving wavelets. The latter half of the book features these developments, the main purpose of *Wavelet Analysis*. The first half of the book attempts to introduce readers to the basic concepts in the study of wavelets.

The authors try to appeal to a wide audience by including a significant amount of expository material. For instance, two sections of the book are titled "Music Notation as a Metaphor for Wavelet Series" and "The Democratization of Arithmetic: Positional Notation for Numbers." An intriguing section relates wavelets to Newton's method; this could be a starting point for an undergraduate research project. A nice example to introduce students to multiresolution analysis can be found in Chapter 3, where the authors describe the "multiresolution representation for a number."

We must caution the reader that despite these fine attributes, this is an advanced work; the authors' expectations of the reader are even higher than those of Blatter and Chui. A footnote early on demonstrates that *Wavelet Analysis* is yet another book belonging to the collection of Fourier-dependent texts: the authors assume that the reader has a complete understanding of the terms "generalized Fourier series," "basis functions," and "variable compact support," which are far beyond what a typical junior mathematics major is likely to have mastered. Furthermore, in the first chapter in the wavelet theory part of the book, the authors venture off into Lie groups and their connection to wavelet matrices. Compounding the difficult nature of the text, there are no problems in the book for students to solve.

To be fair, the book does not claim to be an introduction, but rather states that "this text [is] for upper-level undergraduates and beginning graduate students" and

promises to relate wavelets to "previously known methods in mathematics and engineering." Despite the expository nature of some of the book, the prospective reader should look elsewhere for an introduction to wavelets.

A Primer on Wavelets and their Scientific Applications comes closer than the other books under review to meeting the ideal of a true introductory text. James S. Walker, the author, recognizes "a real need for a simple introduction, a *primer*, which uses only elementary linear algebra and a smidgen of calculus to explain the underlying ideas behind wavelet analysis, and devotes the majority of its pages to explaining how these underlying ideas can be applied to solve significant problems in audio and image processing and in biology and medicine." He begins with an introduction to the Haar wavelets, using only a minimal amount of linear algebra, and uses them to introduce many basic ideas—from averaging and differencing to multiresolution analysis—through concrete examples. The chapter concludes with applications of the Haar wavelets to compressing and denoising audio signals. Subsequent chapters introduce the reader to the Daubechies wavelets, two-dimensional wavelet transforms, the Discrete Fourier Transform (DFT), and wavelet packets. Many applications of wavelets are presented, including compression and denoising of images, edge recognition and enhancement, image recognition, and speech analysis.

There is a lot to like in *A Primer on Wavelets*. With the exception of the chapter on frequency analysis, where the DFT is discussed, and subsequent related material, the only mathematical background necessary is some linear algebra, specifically dot products and matrix operations. The best part of the book is the depth and variety of applications that are discussed, with an emphasis throughout on the accuracy of the method being used.

Despite the excellent overview of applications, the book falls short of being an ideal introduction for students. Other than in the first chapter and the sections on wavelet packets, the book does not work through specific examples in detail. In addition, despite the jacket's claim that "throughout the text are numerous suggestions for computer experiments and exercises," there are *no* problems for the reader to work.

Walker has written an impressive, freely available program FAWAV [9] to demonstrate wavelets in action. This Windows-based package, which "*requires no programming to use*" (author's emphasis), enables the user to process one- and two-dimensional signals and images. The program even contains a basic audio editor for clipping portions of sound files to study.

With a modest amount of experimentation and on-line help, the user is soon able to use FAWAV to plot functions, load two-dimensional grayscale images in a broad range of formats (including PGM), and manipulate audio clips. Through the transform feature, the wavelet enthusiast can choose from a variety of Haar, Daubechies, and Coifman transforms to see the transform and inverse transform of a signal in a step-by-step fashion. Via wavelet series, one can experiment with various levels of thresholding and see the end results of compression and denoising. It is here that the software may be most valuable to students, for through trial and error they can see how well (or poorly) information can be retrieved following compression or denoising (see Figure 2). The program also includes wavelet packets, Fourier transforms, and a varied collection of tools for detailed study of the effectiveness and accuracy of signal compression. FAWAV offers great potential for in-class demonstrations of many applications of wavelets.

Walker's *Primer* is filled with many figures of FAWAV output that illustrate key ideas. These images, like his text, do much to show the powerful results of wavelet

Figure 2. FAWAV output: a noisy *Lena* (Gr 1), a denoised *Lena* (Gr 2), significance map from thresholded transform used in denoising (Gr 3), and the original *Lena* (Gr 4).

analysis. While this convinces the reader that wavelets work, it often leaves one wondering *how* they work. Like *Toolware*, all of FAWAV is pre-coded, so the algorithms are inaccessible. This limits the program's value for introducing the ideas behind the output to students. In addition, although the author claims that the software is designed to enable the reader to "duplicate all of the applications described in this primer", we were unable to do so in a few cases. A collection of sample exercises to introduce the user to FAWAV would be a valuable improvement; the omission of such exercises and supplementary by-hand problems is the work's greatest shortcoming.

Easily the most accessible text among those under consideration, *A Primer on Wavelets and Their Scientific Applications* is an excellent resource book, especially for its overview of applications.

Of the several dozen books on wavelets published in the last ten years, most have endeavored to create comprehensive and rigorous presentations. With the exception of Walker's text, the books discussed in this review belong to this class. While they may have desired to write introductions, most of the authors of these books have fallen into a familiar trap: considering the most general case first (which, for wavelets, involves a treasure trove of Fourier transform theory) and using an occasional specific case as an example. This is fine for a rigorous, abstract reference book, but it is deadly when trying to *introduce* a topic to a broad audience, particularly one including undergraduate students. And though Walker writes for a broader audience of scientists, his book is not ideally suited to study by undergraduates or novices.

The number of requests received at our web site leads us to believe that there remains a real need for a book on the topic that is written at a truly introductory level. Such a text would be geared to individuals who need an entry point to the more technical books and papers, would provide an appropriate amount of detail (via linear algebra) as to how wavelets work, and would appeal to undergraduate students and non-mathematicians. With its beauty, power, and accessibility, the subject deserves a presentation that further widens the growing collection of wavelet enthusiasts.

REFERENCES

1. Edward Aboufadel and Steven Schlicker, Discovering Wavelets website, `http://www.gvsu.edu/mathstat/wavelets.htm`.
2. Jonathan N. Bradley, Christopher M. Brislawn, and Tom Hopper, The FBI wavelet/scalar quantization standard for grey-scale fingerprint image compression, in *Visual Info. Process. II*, volume 1961 of Proc. SPIE, Orlando, FL, 1993, pp. 293–304.
3. Christopher M. Brislawn, Fingerprints go digital, *Notices Amer. Math. Soc.* **42** (1995) 1278–1283.
4. Barry Cipra, Wavelet applications come to the fore, *SIAM News* **26-7**, no. 7, November 1993, also available online at `http://www.siam.org/siamnews/mtc/mtc1193.htm`.
5. Steven Courtney, Information age multiplies uses of mathematics formulas for the future, *The Hartford Courant*, October 21, 1993, p. E1.
6. Ingrid Daubechies, *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, 1992.
7. Peter Schröder, Wavelet image compression: Beating the bandwidth bottleneck, *Wired*, May 1995, p. 78.
8. Gilbert Strang, Wavelet transforms versus Fourier transforms, *Bull. Amer. Math. Soc.* **28** (1993) 288–305.
9. James S. Walker, FAWAV wavelets software, available online at `http://www.crcpress.com/edp/download/fawav/fawav.htm`.

Grand Valley State University, Allendale, MI 49401
aboufade@gvsu.edu, boelkinm@gvsu.edu, schlicks@gvsu.edu

Poincaré and the Three Body Problem. By June Barrow-Green. American Mathematical Society, 1997, 272 pp., $39.

Reviewed by **Daniel Henry Gottlieb**

In a work of impressive scholarship, the author takes us through the history of the n-body problem from Newton to the present. The center of her story is the prize competition in honor of the 60th birthday of King Oscar II of Sweden in 1889. With royal patronage, with the most prestigious mathematicians as judges, and with the momentous mathematical problem of Civilization as a topic, it had captured the attention of the mathematical world. And the winner was... Poincaré... with a manuscript that had a major error!

The paper was due to be published on the King's birthday a few weeks hence, when Poincaré himself discovered the false result. The difficulty of his position was enormous. An error in a paper so highly honored not only would be a great personal embarrassment, but would damage the reputations of the judges and the organizers of the competition as well as ruin the King's birthday.

Poincaré wrote a letter admitting the mistake (surely the most difficult one a mathematician ever had to write), stopped the presses, paid for the printing costs (which exceeded the prize money by 1000 Kroner), and worked feverishly on a new manuscript, which was printed a year later. The letter and the first suppressed manuscript remained hidden in the archives of the Mittag-Leffler Institute and were only recently rediscovered.

The author analyzes the suppressed flawed manuscript along with the published corrected copy. What underlay Poincaré's error "is arguably the first description of chaotic motion within a dynamical system." The author goes into mathematical detail in tracing the influence of this manuscript, and of later ones by Poincaré on the subjects of Differential Equations, Dynamical Systems, and Celestial Mechanics. Since she has a clear way of describing research, this mathematical detail should be interesting for the practitioners of those disciplines. For the rest of us, she tells about the controversy between the dynamical astronomers and Poincaré, the final solution to the three-body problem, the mathematical personalities and politics of the competition, and much else. Her scholarship gives a firm historical base for reflections about what a mathematician really is.

For example, Arthur Jaffe and Frank Quinn's controversial article [**1**] discusses the issue of published results with inadequate proofs. Among the cautionary tales mentioned is Poincaré's discovery of homology. "Poincaré claimed too much, proved too little, and his *reckless* methods could not be imitated. The result was a dead area which had to be sorted out before it could take off." The context of these remarks imputes a kind of dishonesty to Poincaré, and claims it retarded the subject for years. However, in view of Poincaré's letter, we can ask the author of those lines whether it seems to him now that Poincaré was dishonest, and we can inquire if he himself would write such a letter if he were in Poincaré's position. As far as the "damage" done by Poincaré, I point out that some of the greatest mathematicians of the time took fifty years before they finally got homology right, and in the process they fundamentally changed the way we view almost all of Mathematics.

Practitioners of mathematics follow two historical traditions. One stems from the dawn of Civilization, the other arose in the time of the Greeks. In the older tradition, Mathematics is the handmaiden of the Arts, Science, and Industry. In the Greek tradition, Mathematics is the Queen of Knowledge, the only real way to Understand Nature.

By "Understand Nature", I mean understand it in the way that a Mathematician understands Mathematics: Clearly, distinctly, without ambiguity. To paraphrase Galileo: Once one tastes this kind of Knowledge, he can never be satisfied with a less perfect kind. Only a Mathematician can taste this kind of Knowledge. (Here I mean Mathematician in an inclusive sense, as opposed to merely a member of a particular profession.) This kind of Knowledge makes Mathematics the Queen of the Sciences, and she will reign forever.

But Mathematics is also the Handmaiden of the Sciences. It is a collection of tools to solve problems, to obtain answers, to describe and to measure and to name. You use it to build a bridge, to survey the land, or to navigate the sea. It perfected the masterpieces of our great painters and cast the horoscopes of our superstitious ancestors. This tradition is much older than the Greek tradition of Mathematics as a pure kind of knowledge, and for most well-educated people it forms their view of what Mathematics is. Those who hold to this tradition may practice their mathematics with skill, but the mathematics is secondary to other considerations. I call these people Practitioners.

Now among the fascinating things in this story is that a Practitioner named Hugo Gyldén, upon obtaining information about Poincaré's original prize-winning paper, claimed that he had already published all of Poincaré's results. This led to a long controversy between those Practitioners called Dynamical Astronomers, and the Mathematicians. The problem was that the Mathematicians and the Astronomers had different ideas about what convergence of a series means:

> To illustrate how mathematicians and astronomers differed over this question, Poincaré compared the possible interpretations of the following two series
>
> $$\sum \frac{(1000)^n}{n!} \quad \text{and} \quad \sum \frac{n!}{(1000)^n}.$$
>
> He argued that a mathematician would consider the first convergent and the second divergent, while an astronomer would label them the other way round. (p. 156)

This must be the best of the math versus physics jokes, because it is true! Yet Poincaré did not convey any criticism. He merely wanted to explain the difference to eliminate misunderstandings. He understood that truncating a divergent series whose initial terms decrease fast could produce numbers that are useful in practical problems, but he pointed out that such methods should not be used to prove theoretical results. And he observed that for practical computations it really does not matter whether or not the series converges: What is important is to have some idea of the upper bound of the errors involved.

Poincaré always said that he learned a great deal from these Practitioners, including Gyldén. Most of us would react in the spirit of Hermite, who "was not impressed by Gyldén's grasp of analysis, describing Gyldén as a ghost from a bygone age, who had been left behind as the world of analysis transformed about him." It turns out, though, that Poincaré had the right point of view, because in 1909, the Finnish astronomer Karl Sundman completely solved the three-body problem! Given an initial position, he could produce a convergent series giving the positions of the bodies for all times.

Wait a minute, I didn't know that the Three-Body Problem was solved. I'll bet you didn't either. "Sundman's work seems to have been almost forgotten. Why did such an important and long awaited work almost fade into obscurity?" Think about it!

The n-body problem can be thought of as the most fateful problem in all of Mathematics. One might say that the mathematician Galileo Galilei "solved" the one-body problem by assuming as an axiom that a body moves with uniform motion in a straight line. Reasoning with other axioms, he showed that a projectile follows a parabolic path. To objections that no one has seen a body move forever in a straight line, he would say: Let us derive the mathematics, and compare the results to those we actually observe. If their differences can be explained by other effects, then the axiom is reasonable. If there is a disparity, then the axiom should be discarded. Thus Galileo followed in the tradition of Archimedes and used Mathematics to Understand Nature.

The two-body problem was essentially solved by Newton in 1687. Newton laid down his three Laws, the first two adapted from Galileo, plus the fourth Law of Gravity. With these axioms mathematicians could understand the workings of the solar system, and they strove to develop methods of calculating. This stupendous achievement, Newtonian Mechanics, led to an entire reorientation of Western

Culture. The following century, called the Age of Reason, found Mathematical ideas applied in every field as people tried to emulate Newton's clarity.

How could the solution of the three-body problem fade into obscurity? Well, the first reason is that Sundman's series does not converge according to the Practitioners. The second reason is that it is an algorithm conveying little insight: although it is precise, it does not add much to Understanding Nature. The third reason is that the Physicist Albert Einstein, noting some slight inadequacy in Galileo's solution of the one-body problem, propounded a new solution, and thereby became a Mathematician.

I thoroughly enjoyed June Barrow-Green's book. I have written things here I would not have dreamt of saying before reading it. For me, the center of the work is Poincaré's letter; for now we can show the Practitioners, and the World, just what we Mathematicians are.

REFERENCES

1. Arthur Jaffe and Frank Quinn, "Theoretical mathematics": toward a cultural synthesis of mathematics and theoretical physics, *Bull. Amer. Math. Soc.* (*N.S.*) **29** (1993) 1–13.

Purdue University, West Lafayette, IN 47907-1395
gottlieb@math.purdue.edu

From the MONTHLY fifty years ago ...

No teacher would attempt to take a student through a college freshman course in mathematics unless he were sure that the student understood automatically, through long familiarity, the meaning of words like multiplication and addition. But what about words like factor and term? These, to the instructor, are just "as simple," just as "automatic"; they are part of his everyday vocabulary. He uses them in class casually, expecting their meaning to be second nature to anyone who has been through high school algebra.

Unfortunately this is not so. To verify a long-standing suspicion that all was not as it should be with the freshman's mathematical vocabulary, the men in three sections were asked, at the beginning of the college year, to write down their definitions of each of five words: *polynomial*, *quotient*, *term*, *coefficient*, and *factor*. The intention was not, of course, to obtain rigorous or polished definitions. If a man showed that he understood the general meaning of the word, he was given the benefit of the doubt. Yet of 60 men quizzed, 33 did not know what a polynomial was; 11 missed quotient; 43 defined term either incorrectly or so badly that it was impossible to tell whether they knew what it was; 22 went astray on coefficient, including 9 who defined it as an exponent; and 19 were hazy on factor

... MONTHLY **56** (1949) 261

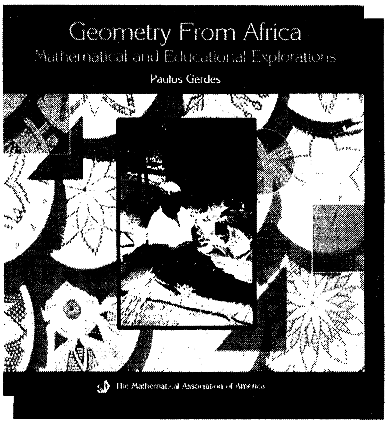
The Mathematical Association of America

Geometry from Africa

Mathematical and Educational Explorations

Paulus Gerdes

Series: Classroom Resource Materials



The peoples of Africa south of the Sahara constitute a vibrant cultural mosaic, extremely rich in its diversity. Among the peoples of the sub-Saharan region, interest in creating and exploring forms and shapes has blossomed in diverse cultural and social contests with such an intensity that with reason it may be said that "Africa Geometrizes".

Gerdes presents examples of geometrical ideas in the work of wood and ivory carvers, potters, painters, weavers, and mat and basket makers. He analyzes geometrical ideas inherent in various crafts and explores possibilities for their educational use. Using as examples African ornaments and artifacts from Senegal to Madagascar, he shows how students may be led to discover the Pythagorean Theorem and to find proofs of it. He also explores connections to Pappus' Theorem, similar right triangles, and Latin and magic squares as well as the geometrical ideas inherent in mat and basket weaving, house building, and wall decoration.

The author presents the geometry of a central African sand drawing tradition--called *sona* in the Chokwe language (predominantly northeast Angola). Through the knowledge of sona, passed from generation to generation via beautiful, often symmetric, designs made in the sand, Gerdes uncovers mathematical ideas and presents examples of how they may be used in teaching mathematics. He underscores the mathematical potential of the sand drawing tradition by developing the geometry of a new type of design/pattern, which he calls Lunda-designs.

Catalog Code: AFR/JR 224 pp., Paperbound, 1999 ISBN 0-88385-715-4 List: $39.95 MAA Member: $31.50

Name__________ Credit Card No.__________
Address__________ Signature__________Exp. Date____/____
City__________ Qty______ Price $______ Amount $______
State______ Zip ______ Shipping and Handling $______
Phone __________ **Catalog Code: AFR/JR** Total $______

Shipping and Handling: Postage and handling are charged as follows: USA orders (shipped via UPS): $2.95 for the first book, and $1.00 for each additional book. **Canadian orders:** $4.50 for the first book and $1.50 for each additional book. Canadian orders will be shipped within 10 days of receipt of order via the fastest available route. We do not ship via UPS into Canada unless the customer specially requests this service. Canadian customers who request UPS shipment will be billed an additional 7% of their total order. **Overseas Orders:** $3.50 per item ordered for books sent surface mail. Airmail service is available at a rate of $7.00 per book. Foreign orders must be paid in US dollars through a US bank or through a New York clearinghouse. **Credit card orders are accepted for all customers. All orders must be prepaid with the exception of books purchased for resale by bookstores and wholesalers.**

Order Via:

Phone: **1 (800) 331.1622**
Fax: **(301) 206.9789**
Mail: **Mathematical Association of America**
PO Box 91112
Washington, DC 20090-1112
Web: **www.maa.org**

THE MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth Street, N.W.
Washington, DC 20036